GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: October 3, 2000, 12:55:34 ; Search time 157.16 Seconds
(without alignments)
550.817 Million cell updates/sec

Title: US-09-065-672-5
Perfect score: 346
Sequence: 1 CTAAGGCGTGCAAACAGAGC.....CTGTCTCTATTTATACAATA 346

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		26	7.5	3200	1	X33947	Human HCMV inducib
2	c	24	6.9	6511	1	Q95493	Human Cdn-2 DNA. N
3		23	6.6	84	1	T25848	Human gene signatu
4	c	22	6.4	840	1	V39298	Human RAD54 nuclei
5		22	6.4	1363	1	T15455	Lung cancer specif
6		22	6.4	2310	1	Q14851	Clone pTB1283 enco
7		22	6.4	2676	1	Q14850	Clone pTB1284 enco
8	c	22	6.4	10380	1	T67164	Human alpha-N-acet
9		22	6.4	13104	1	Q46852	Clone of recombina
10		22	6.4	20303	1	T71699	Human deoxycytidyl
11		22	6.4	26764	1	T71696	Human deoxycytidyl
12	c	22	6.4	35100	1	V20441	Human c-fms oncoge
13		22	6.4	80240	1	V83940	NC-contig derived
14		22	6.4	80595	1	V83939	HC-contig derived
15		21	6.1	158	1	T25057	Human gene signatu
16	c	21	6.1	262	1	T22201	Human gene signatu
17		21	6.1	384	1	Q60667	Human brain Expres
18	c	21	6.1	423	1	Q60666	Human brain Expres
19		21	6.1	1015	1	X30159	Human secreted pro
20		21	6.1	1534	1	T18324	BRCA1 gene 5' tran
21		21	6.1	1534	1	T32611	BRCA1 gene 5' tran
22		21	6.1	3798	1	V36328	Human BRCA1 gene p
23		21	6.1	4009	1	T85827	Human interleukin-
24		21	6.1	11811	1	V83943	Bacterial artifici
25		21	6.1	24025	1	T17455	Mutated BRCA1 geno
26		21	6.1	24025	1	T17515	Mutated BRCA1 geno
27		21	6.1	24026	1	T18325	BRCA1, human breas
28		21	6.1	24026	1	T17512	Mutated BRCA1 geno
29		21	6.1	24026	1	T17513	Mutated BRCA1 geno
30		21	6.1	24026	1	T17514	Mutated BRCA1 geno
31		21	6.1	24026	1	T17516	Mutated BRCA1 geno
32		21	6.1	24026	1	T17517	Mutated BRCA1 geno
33		21	6.1	24026	1	T17518	Mutated BRCA1 geno
34		21	6.1	24026	1	T17519	Mutated BRCA1 geno
35		21	6.1	24026	1	T17521	Mutated BRCA1 geno
36		21	6.1	24026	1	T17522	Mutated BRCA1 geno
37		21	6.1	24026	1	T17523	Mutated BRCA1 geno
38		21	6.1	24026	1	T17524	Mutated BRCA1 geno
39		21	6.1	24026	1	T17526	Mutated BRCA1 geno
40		21	6.1	24026	1	T17527	Mutated BRCA1 geno
41		21	6.1	24026	1	T17528	Mutated BRCA1 geno
42		21	6.1	24026	1	T17529	Mutated BRCA1 geno
43		21	6.1	24026	1	T17530	BRCA1 genomic sequ
44		21	6.1	24026	1	T32612	BRCA1, human breas
45		21	6.1	24029	1	T17520	Mutated BRCA1 geno

ALIGNMENTS

RESULT 1
X33947
ID X33947 standard; DNA; 3200 BP.
AC X33947;
DT 30-JUN-1999 (first entry)
DE Human HCMV inducible gene, SEQ ID NO 21.
KW HCMV inducible gene; cig; human; human cytomegalovirus; interferon;
KW anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
KW drug screening; ds.
OS Homo sapiens.
PN WO9913075-A2.
PD 18-MAR-1999.
PF 08-SEP-1998; U18638.
PR 22-SEP-1997; US-059725.
PR 08-SEP-1997; US-058180.
PA (UYPR-) UNIV PRINCETON.
PI Cong J, Schenk T, Zhu H;
DR WPI; 99-243729/20.
PT New isolated human genes
PS Claim 2; Page 143-147; 184pp; English.
CC This sequence represents a human gene of the invention, that is induced
CC to express by both HCMV and interferon (IFN), designated HCMV-inducible
CC genes (cig or cigs). The invention also relates to genes that are
CC repressed in the presence of HCMV infection, designated HCMV-repressible
CC genes (crg or crgs). The products can be used to obtain agents which can
CC be used for anti-viral therapy, particularly anti-HCMV therapy. They can
CC also be used for the development of drugs that would allow for higher
CC dosage IFN treatments without the concomitant toxicity normally
CC associated with administering high levels of IFN. The products can also
CC be used for detection, diagnosis and drug screening.
SQ Sequence 3200 BP; 972 A; 629 C; 742 G; 857 T;

Query Match 7.5%; Score 26; DB 1; Length 3200;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGGGCAA 314
||||||||||||||||||||||||||
Db 380 CAGGAGTTCCAGACCAGCCTGGGCAA 405

RESULT 2
Q95493/c
ID Q95493 standard; DNA; 6511 BP.
AC Q95493;
DT 21-NOV-1995 (first entry)
DE Human Cdn-2 DNA.
KW Cdn-2; apoptosis modulator; adoptive immunotherapy; therapy; HIV;
KW autoimmune disease; reperfusion injury; hepatitis, osteoporosis;
KW shock; lymphoma; eczema; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 3312. .3947
FT /*tag= a
PN WO9515084-A.

PD 08-JUN-1995.
PF 30-NOV-1994; U13930.
PR 30-NOV-1993; US-160067.
PR 07-OCT-1994; US-320157.
PA (LXRB-) LXR BIOTECHNOLOGY INC.
PI Barr PJ, Kiefer MC;
DR WPI; 95-215106/28.
DR P-PSDB; R77877.
PT New nucleic acid sequences encoding Cdn apoptosis modulators - and
PT related vectors, transformed cells, proteins and antibodies, useful
PT or diagnosis and treatment e.g. of HIV infection, reperfusion injury
PT etc.
PS Claim 6; Fig.5A-H; 66pp; English.
CC Cdn-2 cDNA was isolated from a human placental genomic library
CC using a 950 bp fragment of Cdn-1 cDNA. Expression of Cdn-2
CC in mouse progenitor B-cell FL5.12 cells decreased IL-3-induced
CC apoptosis. The Cdn-2 protein displayed 97% amino acid identity
CC with Cdn-1 (R77876).
SQ Sequence 6511 BP; 1513 A; 1620 C; 1605 G; 1773 T;

Query Match 6.9%; Score 24; DB 1; Length 6511;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGGGC 312
||||||||||||||||||||||||
Db 1393 CAGGAGTTCCAGACCAGCCTGGGC 1370

RESULT 3
T25848
ID T25848 standard; cDNA to mRNA; 84 BP.
AC T25848;
DT 22-OCT-1996 (first entry)
DE Human gene signature HUMGS08078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1942; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 84 BP; 33 A; 17 C; 15 G; 19 T;

Query Match 6.6%; Score 23; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGGG 311
|||||||||||||||||||||||
Db 14 CAGGAGTTCCAGACCAGCCTGGG 36

RESULT 4
V39298/c
ID V39298 standard; cDNA; 840 BP.
AC V39298;
DT 16-SEP-1998 (first entry)
DE Human RAD54 nucleic acid sequence comprising exon 9.
KW Human; RAD54; hRAD54; cancer; xeroderma pigmentosum; Bloom syndrome;
KW Werner's syndrome; ATR-X; diagnosis; detection; SNF2 superfamily;
KW X-linked mental retardation with alpha-thalassemia syndrome; tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN EP-844305-A2.
PD 27-MAY-1998.
PF 10-NOV-1997; 308998.
PR 13-NOV-1996; US-030676.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (UYJE-) UNIV JEFFERSON THOMAS.
PI Croce CM, Fishel RA, Rasio D, Robbins DJ;
DR WPI; 98-274189/25.
PT Human hRAD54 DNA and polypeptide - and agonists, antibodies,
PT antagonists, etc.
PS Claim 1; Page 28; 64pp; English.
CC The present sequence represents a specifically claimed partial nucleic
CC acid sequence encoding human RAD54 (hRAD54). A method for analysing a
CC sample for mutation of DNA encoding hRAD54 has been developed using a
CC DNA sequence of at least 15 and no more than 30 consecutive bases of
CC the DNA sequence encoding hRAD54. hRAD54 is a gene thought to be present
CC in tumours that display allelic imbalance at Ip32, the chromosomal band
CC identified as one of four minimal regions of chromosome 1 deletion in
CC breast carcinomas. hRAD54 is useful for production of proteins, inter
CC alia, that have been identified as novel hRAD54 by homology between the
CC amino acid sequence given in W62186 and known amino acid sequences such
CC as yeast RAD54. hRAD54 proteins are used in the treatment of cancer,
CC including Xeroderma Pigmentosum and Bloom syndrome, Werner's syndromes
CC and X-linked mental retardation with alpha-thalassaemia syndrome and
CC breast cancer. hRAD54 polynucleotides are also useful for detecting
CC complementary nucleotides for use as a diagnostic agent, especially
CC useful for diagnosis of disease or susceptibility to diseases. hRAD54
CC polynucleotide, proteins, agonists and antagonists which are proteins
CC are useful in gene therapy.
SQ Sequence 840 BP; 190 A; 200 C; 221 G; 229 T;

Query Match 6.4%; Score 22; DB 1; Length 840;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 150 CAGGAGTTCCAGACCAGCCTGG 129

RESULT 5
T15455
ID T15455 standard; DNA; 1363 BP.
AC T15455;
DT 23-APR-1996 (first entry)
DE Lung cancer specific antigen HCAVIII promoter region genomic DNA.
KW Non-small cell lung cancer; NSCLC; tumour marker; HCAVIII;
KW carbonic anhydrase; diagnosis; therapy; promoter; DNA probe;
KW fluorescent in situ hybridisation; ds.
OS Homo sapiens.
PN WO9602552-A1.
PD 01-FEB-1996.
PF 19-JUL-1995; U09145.
PR 19-JUL-1994; US-276919.

```
PA   (CYTO-) CYTOCLONAL PHARM INC.
PI   Bollon AP, Torczynski RM;
DR   WPI; 96-105844/11.
PT   Nucleic acid encoding the lung cancer specific antigen HCAVIII -
PT   useful for diagnosis and treatment of non-small cell lung cancer
PS   Claim 53; Page 62-63; 87pp; English.
CC   A genomic clone (T15455) was isolated that constitutes the putative
CC   promoter of the HCAVIII gene (see T15448), and probably contains
CC   transcription regulatory elements directly implicated in expression
CC   of HCAVIII, a cell surface antigen which is highly specific for
CC   non-small cell lung carcinoma and which has features in common with
CC   human carbonic anhydrases. The clone was obtd. by PCR amplification
CC   using a primer pair (T15456-57) based on the putative exon 6 of the
CC   HCAVIII gene. A DNA probe comprising the genomic clone plus
CC   flanking sequences was used for fluorescent in situ hybridisation.
SQ   Sequence   1363 BP;    352 A;    382 C;    369 G;    260 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 1363;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     289 CAGGAGTTCCAGACCAGCCTGG 310
           ||||||||||||||||||||||
Db     554 CAGGAGTTCCAGACCAGCCTGG 575


RESULT   6
Q14851
ID   Q14851 standard; DNA; 2310 BP.
AC   Q14851;
DT   18-FEB-1992  (first entry)
DE   Clone pTB1283 encoding complete FGF receptor.
KW   Human; fibroblast growth factor; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             25. .1983
FT                   /*tag= a
PN   WO9117183-A.
PD   14-NOV-1991.
PF   25-APR-1991; J00557.
PR   27-APR-1990; JP-113146.
PR   31-JUL-1990; JP-204438.
PR   14-SEP-1990; JP-245256.
PR   28-DEC-1990; JP-415801.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
PI   Igarashi K, Senoo M, Watanabe T;
DR   WPI; 91-353723/48.
DR   P-PSDB; R15269.
PT   New mutein(s) of proteins - with fibroblast growth factor
PT   receptor activity, useful for treating multiple endocrine
PT   neoplasia, prostatic hypertrophy, used for diagnosis
PS   Example 3; Fig 8; 88pp; English.
CC   A cDNA library prepared from human cancer cell line Kato III mRNA
CC   was screened with an oligonucleotide corresponding to amino acids
CC   529-541 of chicken basic FGF receptor. Three positive clones were
CC   obtained. One was cloned into pUC118/119 to give pTB1228 (see
CC   Q14848). The complete FGF coding sequence was obtained by ligating
CC   the insert from pTB1228 to the DNA sequence of the plasmid pTB1281
CC   insert which encodes the carboxyl terminus of the FGF receptor from
CC   Glu 533 onwards.
SQ   Sequence   2310 BP;    629 A;    566 C;    636 G;    479 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 2310;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     289 CAGGAGTTCCAGACCAGCCTGG 310
           ||||||||||||||||||||||
Db    2088 CAGGAGTTCCAGACCAGCCTGG 2109


RESULT   7
Q14850
ID   Q14850 standard; DNA; 2676 BP.
AC   Q14850;
DT   18-FEB-1992  (first entry)
DE   Clone pTB1284 encoding complete FGF receptor.
KW   Human; fibroblast growth factor; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             25. .2334
FT                   /*tag= a
PN   WO9117183-A.
PD   14-NOV-1991.
PF   25-APR-1991; J00557.
PR   27-APR-1990; JP-113146.
PR   31-JUL-1990; JP-204438.
PR   14-SEP-1990; JP-245256.
PR   28-DEC-1990; JP-415801.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
PI   Igarashi K, Senoo M, Watanabe T;
DR   WPI; 91-353723/48.
DR   P-PSDB; R15268.
PT   New mutein(s) of proteins - with fibroblast growth factor
PT   receptor activity, useful for treating multiple endocrine
PT   neoplasia, prostatic hypertrophy, used for diagnosis
PS   Example 3; Fig 7; 88pp; English.
CC   A cDNA library prepared from human cancer cell line Kato III mRNA
CC   was screened with an oligonucleotide corresponding to amino acids
CC   529-541 of chicken basic FGF receptor. Three positive clones were
CC   obtained. One was cloned into pUC118/119 to give pTB1229 (see
CC   Q14849). The complete FGF coding sequence was obtained by ligating
CC   the insert from pTB1229 to the DNA sequence of the plasmid pTB1281
CC   insert which encodes the carboxyl terminus of the FGF receptor from
CC   Glu 533 onwards.
SQ   Sequence   2676 BP;    743 A;    645 C;    738 G;    550 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 2676;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     289 CAGGAGTTCCAGACCAGCCTGG 310
           ||||||||||||||||||||||
Db    2439 CAGGAGTTCCAGACCAGCCTGG 2460


RESULT   8
T67164/c
ID   T67164 standard; cDNA; 10380 BP.
AC   T67164;
DT   20-AUG-1997  (first entry)
DE   Human alpha-N-acetylglucosaminidase gene.
KW   Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;
KW   gene therapy; enzyme replacement therapy; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   5'utr           1. .989
FT                   /*tag= a
FT   exon            990. .1372
FT                   /*tag= b
FT                   /number= 1
FT   intron          1373. .2114
FT                   /*tag= c
FT   exon            2115. .2262
FT                   /*tag= d
FT                   /number= 2
FT   intron          2263. .3055
FT                   /*tag= e
FT   intron          3056. .3202
FT                   /*tag= f
FT                   /number= 3
FT   intron          3203. .3386
FT                   /*tag= g
```

```
FT   exon            3387. .3472
FT                   /*tag= h
FT                   /number= 4
FT   intron          3473. .5666
FT                   /*tag= i
FT   exon            5667. .5923
FT                   /*tag= j
FT                   /number= 5
FT   intron          5924. .7744
FT                   /*tag= k
FT   exon            7745. .8955
FT                   /*tag= l
FT                   /number= 6
FT   3'utr           8966. .10380
FT                   /*tag= m
PN   WO9719177-A1.
PD   29-MAY-1997.
PF   22-NOV-1996; AU0747.
PR   23-NOV-1995; AU-006748.
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI   Anson DS, Blanch L, Hopwood JJ, Scott H, Weber B;
DR   WPI; 97-298114/27.
DR   P-PSDB; W18017.
PT   Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -
PT   used for the diagnosis and treatment of mucopolysaccharidosis type
PT   IIIB, also used in gene therapy
PS   Claim 8; Page 54-61; 79pp; English.
CC   A genomic DNA molecule (T67164) includes 6 exons that code for
CC   human alpha-N-acetylglucosaminidase (W18017), an enzyme that can
CC   hydrolyse the terminal alpha-N-acetylglucosamine residues at the
CC   non-reducing terminus of fragments of heparan sulphate and heparin.
CC   It was isolated by hybridisation of a human chromosome 17 library.
CC   A cDNA clone (T67163) coding for the enzyme has also been isolated.
CC   The isolated gene or cDNA, and primers/probes based on them or
CC   their complementary strands, can be used to investigate, diagnose
CC   and treat alpha-N-acetylglucosaminidase deficiency, for example in
CC   patients suffering from mucopolysaccharidosis type IIIB.
CC   Administration is by oral, i.v., i.p., enzyme replacement therapy,
CC   gene therapy or other routes.
SQ   Sequence   10380 BP;    2210 A;    2953 C;    2851 G;    2366 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 10380;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
D     439 CAGGAGTTCCAGACCAGCCTGG 7418


RESULT   9
Q46852/c
ID   Q46852 standard; DNA; 13104 BP.
AC   Q46852;
DT   26-JAN-1994  (first entry)
DE   Clone of recombinant human kappa casein gene fragment.
KW   Casein; supplement; milk; pharmaceutical; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1. .8834
FT                   /*tag= a
FT   exon            8835. .8867
FT                   /*tag= b
FT   intron          8868. .10014
FT                   /*tag= c
FT   exon            10015. .10510
FT                   /*tag= d
FT   intron          10511. .12277
FT                   /*tag= e
FT   exon            12278. .12443
FT                   /*tag= f
PN   WO9315196-A.
PD   05-AUG-1993.
PF   25-JAN-1993; DK00024.
PR   23-JAN-1992; DK-000088.
PA   (SYMB-) SYMBICOM AB.
PI   Bergstroem S, Hansson L, Hernell O, Stroemqvist M;
PI   Toernell J;
DR   WPI; 93-258675/32.
PT   DNA encoding human kappa-casein - used for obtaining recombinant
PT   polypeptide(s) for use as nutrient supplements, partic. in infant
PT   formulae
PS   Example 4; Page 92-99; 110pp; English.
CC   The recombinant human kappa casein is produced in high yields by
CC   means of either a eukaryotic or prokaryotic expression system.  It
CC   is used as a nutrient supplement in milk based products to provide a
CC   substantial improvement of the nutritional and biological value of
CC   the formulae, making it closer in similarity to human milk.  It can
CC   also be used as a pharmaceutical.
SQ   Sequence   13104 BP;    4256 A;    2497 C;    2397 G;    3953 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 13104;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db    327 CAGGAGTTCCAGACCAGCCTGG 306


RESULT  10
T71699
ID   T71699 standard; DNA; 20303 BP.
AC   T71699;
DT   20-AUG-1997  (first entry)
DE   Human deoxycytidylate deaminase intron 2 encoding DNA.
KW   Recombinant deaminase; dCMP; ds.
OS   Homo sapiens.
PN   US5622851-A.
PD   22-APR-1997.
PF   10-JAN-1995; 370975.
PR   10-JAN-1995; US-370975.
PA   (HEAL-) HEALTH RES INC.
PI   Maley F,  Maley GR,  Weiner KXB;
DR   WPI; 97-244391/22.
PT   DNA encoding human deoxycytidylate deaminase - for production of
PT   recombinant deaminase
PS   Claim 2; Column 83-100; 58pp; English.
CC   The present sequence encodes the human deoxycytidylate (dCMP)
CC   deaminase intron 2, which comprises 20303 base pairs from nucleotides
CC   1964-22266 of the dCMP deaminase sense strand. The dCMP deaminase gene
CC   contains a 5' untranslated region (including the promoter), 5 exons,
CC   4 introns and a 3' untranslated region (including the stop signals).
CC   The gene can be used to produce recombinant dCMP deaminase, which can
CC   be used to convert dCMP to dUMP. Also, the dCMP gene can be altered
CC   (removed or mutated) to alter DNA replication in cells, which may lead
CC   to mutagenesis.
SQ   Sequence   20303 BP;    5454 A;    4115 C;    5052 G;    5682 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 20303;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  15284 CAGGAGTTCCAGACCAGCCTGG 15305


RESULT  11
T71696
ID   T71696 standard; DNA; 26764 BP.
AC   T71696;
DT   20-AUG-1997  (first entry)
```

```
DE   Human deoxycytidylate deaminase gene.
KW   Recombinant deaminase; dCMP; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    1. .1317
FT                   /*tag= a
FT                   /note= "5' untranslated region, including promotor"
FT   exon            1318. .1425
FT                   /*tag= b
FT                   /number= 1
FT   intron          1426. .1827
FT                   /*tag= c
FT                   /number= 1
FT   exon            1828. .1963
FT                   /*tag= d
FT                   /number= 2
FT   intron          1964. .22266
FT                   /*tag= e
FT                   /number= 2
FT   exon            22267. .22383
F                    /*tag= f
F                    /number= 3
FT   intron          22384. .23740
FT                   /*tag= g
FT                   /number= 3
FT   exon            23741. .23837
FT                   /*tag= h
FT                   /number= 4
FT   intron          23838. .25391
FT                   /*tag= i
FT                   /number= 4
FT   exon            25392. .25467
FT                   /*tag= j
FT                   /number= 5
FT   misc_feature    25468. .26764
FT                   /*tag= k
FT                   /note= "3' untranslated region"
PN   US5622851-A.
PD   22-APR-1997.
PF   10-JAN-1995; 370975.
PR   10-JAN-1995; US-370975.
PA   (HEAL-) HEALTH RES INC.
PI   Maley F, Maley GR, Weiner KXB;
DR   WPI; 97-244391/22.
DR   P-PSDB; W18205.
PT   DNA encoding human deoxycytidylate deaminase - for production of
PT   recombinant deaminase
PS   Claim 3; Column 55-78; 58pp; English.
C    The present sequence encodes the human deoxycytidylate (dCMP)
C    deaminase gene, which contains a 5' untranslated region (including
CC   the promoter), 5 exons, 4 introns and a 3' untranslated region
CC   (including the stop signals). The gene can be used to produce
CC   recombinant dCMP deaminase, which can be used to convert dCMP to dUMP.
CC   Also, the dCMP gene can be altered (removed or mutated) to alter DNA
CC   replication in cells, which may lead to mutagenesis.
SQ   Sequence   26764 BP;    7079 A;    5521 C;    6539 G;    7625 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 26764;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  17247 CAGGAGTTCCAGACCAGCCTGG 17268


RESULT  12
V20441/c
ID   V20441 standard; DNA; 35100 BP.
AC   V20441;
DT   17-JUN-1998  (first entry)
DE   Human c-fms oncogene.
KW   Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
KW   cancer; antisense oligonucleotide; c-fms; ds.
OS   Homo sapiens.
PN   US5734039-A.
PD   31-MAR-1998.
PF   15-SEP-1994; 306691.
PR   15-SEP-1994; US-306691.
PA   (UYJE-) UNIV JEFFERSON THOMAS.
PI   Calabretta B, Skorski T;
DR   WPI; 98-229882/20.
PT   Anticancer composition comprising two anti-sense oligo:nucleotide(s)
PT    - targetting cytoplasmic and nuclear oncogene(s)
PS   Claim 1; Column 59-90; 92pp; English.
CC   The present sequence represents an oncogene from the present invention.
CC   The present invention describes a composition which comprises two
CC   antisense oligonucleotides. The first oligonucleotide is specific for a
CC   cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC   c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
CC   c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC   proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC   c-rel, c-vav, c-ski, c-spi, cyclin D1, PML/RAR alpha , AML1/MTG8,
CC   E2A/prl and ALL-1/AF-4. The composition is used for treating cancer.
CC   The combination of antisense oligonucleotides has synergistically
CC   enhanced ability to inhibit growth of cancer cells.
SQ   Sequence   35100 BP;    8474 A;    8597 C;    9682 G;    8347 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 35100;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  33551 CAGGAGTTCCAGACCAGCCTGG 33530


RESULT  13
V83940
ID   V83940 standard; DNA; 80240 BP.
AC   V83940;
DT   03-MAR-1999  (first entry)
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS   Homo sapiens.
PN   WO9851790-A1.
PD   19-NOV-1998.
PF   13-MAY-1998; AU0352.
PR   26-AUG-1997; AU-008791.
PR   13-MAY-1997; AU-006784.
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
PI   Cancilla MR, Choo K, Du Sart D;
DR   WPI; 99-009773/01.
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy
PS   Claim 9; Fig 16A; 540pp; English.
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains
CC   an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in
CC   a compatible cell, of replicating, acting as extra-chromosomal element
CC   and segregating during cell division. The sequences can be used to
CC   construct artificial chromosomes for use in gene therapy comprising a
CC   replicable, segregating nucleic acid that confers a specific phenotype
CC   on cells. Human artificial chromosomes can propagate in human cells and
```

```
CC   carry large amounts of DNA (e.g. therapeutic genes), and, being
CC   extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC   are also useful for generation of transgenic plants and animals, in
CC   production of proteins and to make diagnostic reagents, e.g. for
CC   expression of cytokines, receptors and growth factors, or to increase
CC   the copy number of a gene in a cell. The constructs may also be
CC   used for functional and structural analysis of chromosomes.
SQ   Sequence   80240 BP;    23102 A;    16537 C;    16747 G;    23846 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 80240;
  Best Local Similarity   100.0%;  Pred. No. 0.023;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  27312 CAGGAGTTCCAGACCAGCCTGG 27333


RESULT  14
V[illegible]9
ID   V83939 standard; DNA; 80595 BP.
AC   V83939;
DT   03-MAR-1999  (first entry)
DE   HC-contig derived from normal human chromosome 10q25.2 region.
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS   Homo sapiens.
PN   WO9851790-A1.
PD   19-NOV-1998.
PF   13-MAY-1998; AU0352.
PR   26-AUG-1997; AU-008791.
PR   13-MAY-1997; AU-006784.
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
PI   Cancilla MR, Choo K, Du Sart D;
DR   WPI; 99-009773/01.
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy
PS   Claim 8; Fig 6; 540pp; English.
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in
CC   a compatible cell, of replicating, acting as extra-chromosomal element
CC   and segregating during cell division. The sequences can be used to
CC   construct artificial chromosomes for use in gene therapy comprising a
CC   replicable, segregating nucleic acid that confers a specific phenotype
CC   on cells. Human artificial chromosomes can propagate in human cells and
CC   carry large amounts of DNA (e.g. therapeutic genes), and, being
CC   extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC   are also useful for generation of transgenic plants and animals, in
CC   production of proteins and to make diagnostic reagents, e.g. for
CC   expression of cytokines, receptors and growth factors, or to increase
CC   the copy number of a gene in a cell. The constructs may also be
CC   used for functional and structural analysis of chromosomes.
SQ   Sequence   80595 BP;    23183 A;    16613 C;    16824 G;    23975 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 80595;
  Best Local Similarity   100.0%;  Pred. No. 0.023;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  27572 CAGGAGTTCCAGACCAGCCTGG 27593


RESULT  15
T25057
ID   T25057 standard; cDNA to mRNA; 158 BP.
AC   T25057;
DT   11-NOV-1996  (first entry)
DE   Human gene signature HUMGS07188.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1759; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   158 BP;    46 A;    35 C;    44 G;    30 T;


  Query Match             6.1%;  Score 21;  DB 1;  Length 158;
  Best Local Similarity   100.0%;  Pred. No. 0.065;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    258 GGGAGGCCGAGGCAGGAAGAT 278
          |||||||||||||||||||||
Db    119 GGGAGGCCGAGGCAGGAAGAT 139


Search completed: October  3, 2000, 12:56:19
Job time: 7379 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: October 3, 2000, 14:37:34 ; Search time 114.21 Seconds
(without alignments)
604.614 Million cell updates/sec

Title: US-09-065-672-4_COPY_1_276
Perfect score: 276
Sequence: 1 CTAAGGCGTGCAAACAGAGC.....CGGGAGGCCGAGGCAGGAAG 276

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

W[redacted]size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19	6.9	158	1	T25057	Human gene signatu
c 2	19	6.9	540	1	V87076	EST clone BJ66. Ne
3	19	6.9	702	1	X22242	Human secreted pro
c 4	18	6.5	2703	1	N90541	DNA encoding N-alp
c 5	18	6.5	2724	1	Q12226	N-alpha-acetyltran
c 6	17	6.2	332	1	V90303	EST clone DK113. N
c 7	17	6.2	338	1	Q59619	Human brain Expres
[redacted] 8	17	6.2	406	1	Q60129	Human brain Expres
[redacted] 9	17	6.2	541	1	V90043	EST clone CW1510.
[redacted] 10	17	6.2	632	1	V88129	EST clone FY354. N
11	17	6.2	688	1	T72060	Sequence flanking
12	17	6.2	688	1	T43940	Sequence flanking
13	17	6.2	2351	1	X30406	DNA encoding a hum
14	17	6.2	3523	1	V11854	Human Duffy genomi
15	17	6.2	3523	1	V27017	Homo sapiens DNA f
c 16	17	6.2	7146	1	V38933	Nucleic acid seque
17	17	6.2	11288	1	Q90512	CEA clone HindIII-
c 18	17	6.2	13585	1	T11549	Tumour rejection a
19	17	6.2	14556	1	Q90511	CEA genomic clone.
20	17	6.2	14557	1	X13304	Enterococcus faeca
21	17	6.2	15056	1	V52967	Carcinoembryonic a
c 22	17	6.2	20199	1	V52139	Streptococcus pneu
c 23	17	6.2	22481	1	T11658	PEDF full length s
24	17	6.2	28720	1	V49655	Human SC3 DNA. Pro
25	17	6.2	110000	1	X20248_01	Continuation (2 of
26	17	6.2	235033	1	V57926	Hereditary haemoch
27	17	6.2	237326	1	V57903	Hereditary haemoch
c 28	17	6.2	237326	1	V57903	Hereditary haemoch
c 29	16	5.8	20	1	V69963	Human c-fos protei
c 30	16	5.8	134	1	T24389	Human gene signatu
c 31	16	5.8	155	1	V75820	Staphylococcus aur
c 32	16	5.8	164	1	T25606	Human gene signatu
33	16	5.8	180	1	X03464	Intron 13 sequence
c 34	16	5.8	258	1	Q61087	Human brain Expres
c 35	16	5.8	260	1	T22202	Human gene signatu
c 36	16	5.8	269	1	Q60189	Human brain Expres
c 37	16	5.8	271	1	T26324	Human gene signatu
c 38	16	5.8	284	1	V88777	EST clone HO107. N
c 39	16	5.8	290	1	Q61371	Human brain Expres
c 40	16	5.8	303	1	V88486	EST clone EN10. Ne
c 41	16	5.8	341	1	Q59782	Human brain Expres
c 42	16	5.8	342	1	T26195	Human gene signatu
c 43	16	5.8	350	1	T26581	Human gene signatu
44	16	5.8	358	1	Q60394	Human brain Expres
c 45	16	5.8	383	1	Q61360	Human brain Expres

ALIGNMENTS

RESULT 1
T25057
ID T25057 standard; cDNA to mRNA; 158 BP.
AC T25057;
DT 11-NOV-1996 (first entry)
DE Human gene signature HUMGS07188.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1759; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 158 BP; 46 A; 35 C; 44 G; 30 T;

Query Match 6.9%; Score 19; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGAAG 276
||||||||||||||||||||
Db 119 GGGAGGCCGAGGCAGGAAG 137

RESULT 2
V87076/c
ID V87076 standard; cDNA; 540 BP.
AC V87076;
DT 27-APR-1999 (first entry)
DE EST clone BJ66.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845435-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06954.
PR 10-APR-1997; US-835913.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070076/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 444; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 540 BP; 116 A; 150 C; 127 G; 147 T;

Query Match 6.9%; Score 19; DB 1; Length 540;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 255 TTCGGGAGGCCGAGGCAGG 273
||||||||||||||||||||
Db 255 TTCGGGAGGCCGAGGCAGG 237

RESULT 3
X22242
ID X22242 standard; DNA; 702 BP.
AC X22242;
DT 18-MAY-1999 (first entry)
DE Human secreted protein gene 32 clone HJABC16.
KW Human; secreted protein; gene therapy; protein therapy; cancer; weight;
KW tumour; chromosome mapping; forensic; haematological disease; allergy;
KW inflammation; cell proliferation; viral infection; wound healing;
KW modulation; appetite; behaviour; food additive; preservative; ss.
OS Homo sapiens.
PN WO9903990-A1.
PD 28-JAN-1999.
PF 15-JUL-1998; U14613.
PR 18-AUG-1997; US-056361.
PR 16-JUL-1997; US-052661.
PR 16-JUL-1997; US-052870.
PR 16-JUL-1997; US-052871.
PR 16-JUL-1997; US-052872.
PR 16-JUL-1997; US-052873.
PR 16-JUL-1997; US-052874.
PR 16-JUL-1997; US-052875.
PR 22-JUL-1997; US-053440.
PR 22-JUL-1997; US-053441.
PR 22-JUL-1997; US-053442.
PR 18-AUG-1997; US-055683.
PR 18-AUG-1997; US-055724.
PR 18-AUG-1997; US-055725.
PR 18-AUG-1997; US-055726.
PR 18-AUG-1997; US-055946.
PR 18-AUG-1997; US-055952.
PR 18-AUG-1997; US-055985.
PR 18-AUG-1997; US-055989.
PR 18-AUG-1997; US-056359.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Duan R, Feng P, Ferrie AM, Florence KA, Fouad J,
PI Greene JM, Hu J, Ni J, Rosen CA, Ruben SM, Young PE,
PI Yu G;
DR WPI; 99-132234/11.
DR P-PSDB; Y01414.
PT New nucleic acids encoding secreted human proteins - potentially
PT useful for treating and diagnosing diseases and identifying specific
PT binding agents
PS Claim 4; Page 185-186; 251pp; English.
CC The invention relates to nucleic acid sequences (X22211 to X22282)
CC encoding human secreted proteins (Y01383 to Y01454). The secreted protein
CC gene sequences are deposited with the ATCC under deposit number ATCC
CC 209138, 209139 or 209141. Host cells containing vectors comprising the
CC nucleic acid sequences are used for the recombinant expression of the
CC secreted proteins. The polynucleotide and amino acid sequences are useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. Pathological conditions can be also diagnosed by
CC determining the amount of the new polypeptides in a sample or by the
CC presence of mutations in the new polynucleotides. The nucleic acid
CC sequences, or its fragments, are useful for chromosome identification and
CC mapping; as antisense and triplex-forming therapeutics; in gene therapy;
CC for (forensic) identification of individuals; as molecular weight
CC markers; to identify related sequences or specific mRNA; in preparation
CC of oligomers and to raise anti-DNA antibodies. Antibodies are useful as
CC immunoassay reagents (including for in vivo imaging) and therapeutically
CC to inhibit or activate particular polypeptides. A very wide range of
CC disorders may be treated with the polynucleotide and polypeptide
CC sequences, e.g. autoimmune or haematological diseases, allergy,
CC inflammation, cancer or other forms of cell proliferation, viral or other
CC infections. The sequences may also be useful in wound healing, to
CC modulate differentiation of embryonic stem cells, to modulate weight,
CC appetite, behaviour etc. and as food additive or preservative. The
CC present sequence represents a gene encoding a human secreted protein
CC (see descriptor line for gene number and clone identification).
SQ Sequence 702 BP; 187 A; 154 C; 174 G; 183 T;

Query Match 6.9%; Score 19; DB 1; Length 702;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 255 TTCGGGAGGCCGAGGCAGG 273
||||||||||||||||||||
Db 441 TTCGGGAGGCCGAGGCAGG 459

RESULT 4
N90541/c
ID N90541 standard; recombinant DNA; 2703 BP.
AC N90541;
DT 28-NOV-1989 (first entry)
DE DNA encoding N-alpha-acetyl transferase.
KW N-alpha-acetyl transferase; herbicide resistance;
KW protein N-acetylation.
FH Key Location/Qualifiers
FT misc_feature 272. .335
FT /*tag= a
FT misc_feature 338. .392
FT /*tag= b
FT misc_feature 479. .515
FT /*tag= c
FT misc_feature 542. .566
FT /*tag= d
FT misc_feature 971. .989
FT /*tag= e
FT misc_feature 1007. .1049
FT /*tag= f
FT misc_feature 1061. .1085

```
FT                   /*tag= g
FT   misc_feature    1088. .1130
FT                   /*tag= h
FT   misc_feature    1481. .1505
FT                   /*tag= i
FT   misc_feature    1829. .1856
FT                   /*tag= j
FT   misc_feature    1862. .1885
FT                   /*tag= k
FT   misc_feature    1909. .1934
FT                   /*tag= l
FT   misc_feature    2072. .2117
FT                   /*tag= m
FT   misc_feature    2123. .2183
FT                   /*tag= n
PN   WO8907138-A.
PD   10-AUG-1989.
PF   07-FEB-1989; U00471.
PR   08-FEB-1988; US-153361.
PR   14-DEC-1988; US-284344.
PA   (GEHO) The General Hospital Corporation.
PI   Smith JA, Lee FJS;
DR   WPI; 89-249008/34.
DR   P-PSDB; P91070.
PT   New pure N-alpha-acetyl transferase and DNA encoding it - catalysing
PT   acetylation of proteins and peptides, eg to stabilise pharmaceuticals
PT   or induce herbicide resistance in plants.
PS   Claim 8; Page 50; fig 12b-e; 72pp; English.
CC   DNA encodes N-alpha-actyl transferase, used for catalysing N-acetylation
CC   of peptides/proteins, eg to stabilise pharmaceuticals or to induce
CC   herbicide resistance in plants. Features a - n are fragments resulting
CC   from exonuclease III deletion. See also P91070.
SQ   Sequence   2703 BP;   943 A;   489 C;   530 G;   741 T;


  Query Match             6.5%;  Score 18;  DB 1;  Length 2703;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db    744 TGCTTTTCTCTCTAATAA 727


RESULT   5
Q12226/c
ID   Q12226 standard; cDNA; 2724 BP.
AC   Q12226;
DT   02-AUG-1991  (first entry)
DE   N-alpha-acetyltransferase AAA1 gene.
KW   N-alpha-acetyltransferase; amino acid sequencing; AAA1 gene; ss.
OS   Saccharomyces cerivisiae.
FH   Key             Location/Qualifiers
FT   cds             22. .2583
FT                   /*tag= a
FT                   /product= N-alpha-acetyltransferase
PN   WO9106673-A.
PD   16-MAY-1991.
PF   15-OCT-1990; U05883.
PR   25-OCT-1989; US-426381.
PA   (GEHO-) GEN HOSPITAL CORP.
PI   Smith JA, Lee FJS;
DR   WPI; 91-164219/22.
DR   P-PSDB; R12042.
PT   Mutant N-alpha-acetyl-transferase - produced from Saccharomyces
PT   cerevisiae for use in amino acid sequence determn.
PS   Disclosure; Fig 1; 77pp; English.
CC   The AAA1 gene is located on chromosome IV and is positioned
CC   adjacent to the 5' flanking sequence of the SIR2 gene.
CC   Cells contg. a mutated AAA1 gene lack N-alpha-acetyltransferase
CC   activity and are used to express, in vitro a recombinant protein or
CC   peptide lacking an acetyl gp. at the alpha-amino gp. or to produce
CC   heterologous proteins. The proteins produced have altered
CC   N-alpha-acetylation characteristics, e.g. increased or decreased
CC   substrate specificity and thermal stability. The amino acid
CC   sequence of such proteins and peptides can be sequenced.
SQ   Sequence   2724 BP;   953 A;   491 C;   533 G;   747 T;


  Query Match             6.5%;  Score 18;  DB 1;  Length 2724;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db    765 TGCTTTTCTCTCTAATAA 748


RESULT   6
V90303/c
ID   V90303 standard; cDNA; 332 BP.
AC   V90303;
DT   15-FEB-1999  (first entry)
DE   EST clone DK113.
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
OS   Homo sapiens.
PN   WO9845436-A2.
PD   15-OCT-1998.
PF   10-APR-1998; U06955.
PR   10-APR-1997; US-838821.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070077/06.
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
PS   Claim 1; Page 497; 618pp; English.
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
SQ   Sequence   332 BP;   71 A;   83 C;   84 G;   94 T;


  Query Match             6.2%;  Score 17;  DB 1;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    258 GGGAGGCCGAGGCAGGA 274
          |||||||||||||||||
Db    285 GGGAGGCCGAGGCAGGA 269


RESULT   7
Q59619/c
ID   Q59619 standard; cDNA; 338 BP.
AC   Q59619;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST01488.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
```

PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI; 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 204; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST01488 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also Q59041-Q61440.
SQ Sequence 338 BP; 77 A; 106 C; 67 G; 87 T;

Query Match 6.2%; Score 17; DB 1; Length 338;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGA 274
||||||||||||||||
Db 163 GGGAGGCCGAGGCAGGA 147

RESULT 8
Q60129/c
ID Q60129 standard; DNA; 406 BP.
AC Q60129;
DT 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST02116.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI; 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 286; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST02116 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also Q59041-Q61440.
SQ Sequence 406 BP; 74 A; 97 C; 110 G; 124 T;

Query Match 6.2%; Score 17; DB 1; Length 406;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 253 ATTTCGGGAGGCCGAGG 269
|||||||||||||||||
Db 37 ATTTCGGGAGGCCGAGG 21

RESULT 9
V90043/c
ID V90043 standard; cDNA; 541 BP.
AC V90043;
DT 15-FEB-1999 (first entry)
DE EST clone CW1510.
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN WO9845436-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06955.
PR 10-APR-1997; US-838821.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070077/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1; Page 413; 618pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
SQ Sequence 541 BP; 108 A; 192 C; 101 G; 140 T;

Query Match 6.2%; Score 17; DB 1; Length 541;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGA 274
||||||||||||||||
Db 149 GGGAGGCCGAGGCAGGA 133

RESULT 10
V88129/c
ID V88129 standard; cDNA; 632 BP.
AC V88129;
DT 12-FEB-1999 (first entry)
DE EST clone FY354.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845437-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06956.
PR 10-APR-1997; US-837312.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070078/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 292; 641pp; English.
CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for

CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 632 BP; 140 A; 186 C; 145 G; 161 T;

Query Match 6.2%; Score 17; DB 1; Length 632;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 256 TCGGGAGGCCGAGGCAG 272
||||||||||||||||
Db 38 TCGGGAGGCCGAGGCAG 22

RESULT 11
T72060
ID T72060 standard; DNA; 688 BP.
AC T72060;
DT 18-AUG-1997 (first entry)
DE Sequence flanking marker 950-2 in HH region of chromosome 6p2.1.
KW Primer; polymerase chain reaction; amplify; hereditary haemochromatosis;
KW HH; mutation; HH-associated allele; base-pair polymorphism; HHP-1;
KW HHP-19; HHP-29; microsatellite repeat allele; genetic marker;
KW interferon treatment; hepatitis C infection; ss.
OS Synthetic.
PN WO9635802-A1.
PD 14-NOV-1996.
PF 06-MAY-1996; U06352.
PR 08-MAY-1995; US-436074.
PR 15-NOV-1995; US-559302.
PR 09-FEB-1996; US-599252.
PA (MERC-) MERCATOR GENETICS INC.
PI Drayna DT, Feder JN, Gnirke A, Kimmel BE, Thomas WJ;
PI Wolff RK;
DR WPI; 96-518690/51.
PT Determn. of the common hereditary haemochromatosis gene mutation -
PT using primers based on novel microsatellite repeat flanking
PT sequences or on base-pair polymorphisms HHP-1, HHP-19 or HHP-29
PS Claim 24; Fig 1P; 67pp; English.
CC The sequences given in T72045-67 represent portions of the genome
CC surrounding several markers of the invention. The markers were
CC identified using the series of primer pairs given in T71973-2044
CC which were used to determine the presence or absence of the common
CC hereditary haemochromatosis (HH) gene mutation in an individual. The
CC method comprised assessing genomic DNA from an individual for the
CC presence or absence of the HH-associated allele of the single base-pair
CC polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional
CC marker comprising the following microsatellite repeat alleles of group
CC A and optionally of group B:
CC Group A: 19D9, 18B4, 1A2, 1E4, 24E2, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,
CC 731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,
CC 950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;
CC Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.
CC The absence of the genotype indicates the likelihood of the presence of
CC the HH mutation. Knowledge of the new genetic markers allows the
CC definition of genotypes characteristic of heterozygous carriers and
CC homozygotes having a HH mutation in their genomic DNA. The potential for
CC HH in an individual interferes with the effectiveness of interferon
CC treatment for hepatitis C infection. By diagnosing this potential, the
CC responsiveness of interferon treatment may be evaluated.
SQ Sequence 688 BP; 213 A; 155 C; 137 G; 171 T;

Query Match 6.2%; Score 17; DB 1; Length 688;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGA 274
|||||||||||||||||
Db 116 GGGAGGCCGAGGCAGGA 132

RESULT 12
T43940
ID T43940 standard; DNA; 688 BP.
AC T43940;
DT 18-AUG-1997 (first entry)
DE Sequence flanking marker 950-2 in HH region of chromosome 6p2.1.
KW Primer; polymerase chain reaction; amplify; hereditary haemochromatosis;
KW HH; mutation; HH-associated allele; base-pair polymorphism; HHP-1;
KW HHP-19; HHP-29; microsatellite repeat allele; genetic marker;
KW interferon treatment; hepatitis C infection; ss.
OS Synthetic.
PN WO9635803-A1.
PD 14-NOV-1996.
PF 08-MAY-1996; U06583.
PR 08-MAY-1995; US-436074.
PR 15-NOV-1995; US-559302.
PR 09-FEB-1996; US-599252.
PA (MERC-) MERCATOR GENETICS INC.
PI Drayna DT, Feder JN, Gnirke A, Kimmel BE, Thomas WJ;
PI Wolff RK;
DR WPI; 96-518691/51.
PT Diagnosing and genotyping of hereditary haemochromatosis (HH) -
PT using primers to detect specific polymorphisms of the HH gene on
PT chromosome 6p2.1 or novel microsatellite markers
PS Claim 24; Fig 1P; 67pp; English.
CC The sequences given in T43925-55 represent portions of the genome
CC surrounding several markers of the invention. The markers were
CC identified using the series of primer pairs given in T71901-72
CC which were used to determine the presence or absence of the common
CC hereditary haemochromatosis (HH) gene mutation in an individual. The
CC method comprised assessing genomic DNA from an individual for the
CC presence or absence of the HH-associated allele of the single base-pair
CC polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional
CC marker comprising the following microsatellite repeat alleles of group
CC A and optionally of group B:
CC Group A: 19D9, 18B4, 1A2, 1E4, 24E2, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,
CC 731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,
CC 950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;
CC Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.
CC The absence of the genotype indicates the likelihood of the presence of
CC the HH mutation. Knowledge of the new genetic markers allows the
CC definition of genotypes characteristic of heterozygous carriers and
CC homozygotes having a HH mutation in their genomic DNA. The potential for
CC HH in an individual interferes with the effectiveness of interferon
CC treatment for hepatitis C infection. By diagnosing this potential, the
CC responsiveness of interferon treatment may be evaluated.
SQ Sequence 688 BP; 213 A; 155 C; 137 G; 171 T;

Query Match 6.2%; Score 17; DB 1; Length 688;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGA 274
|||||||||||||||||
Db 116 GGGAGGCCGAGGCAGGA 132

RESULT 13
X30406
ID X30406 standard; DNA; 2351 BP.
AC X30406;
DT 14-MAY-1999 (first entry)
DE DNA encoding a human secreted protein.
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;

KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
OS Homo sapiens.
PN WO9907891-A1.
PD 18-FEB-1999.
PF 04-AUG-1998; U16235.
PR 19-AUG-1997; US-056732.
PR 05-AUG-1997; US-054798.
PR 05-AUG-1997; US-054803.
PR 05-AUG-1997; US-054804.
PR 05-AUG-1997; US-054806.
PR 05-AUG-1997; US-054807.
PR 05-AUG-1997; US-054808.
PR 05-AUG-1997; US-054809.
PR 05-AUG-1997; US-055309.
PR 05-AUG-1997; US-055310.
PR 05-AUG-1997; US-055312.
PR 05-AUG-1997; US-055386.
PR 05-AUG-1997; US-055311.
PR 18-AUG-1997; US-055970.
PR 18-AUG-1997; US-055986.
PR 19-AUG-1997; US-056365.
PR 19-AUG-1997; US-056366.
PR 19-AUG-1997; US-056557.
PR 19-AUG-1997; US-056370.
PR 19-AUG-1997; US-056371.
PR 19-AUG-1997; US-056563.
PR 19-AUG-1997; US-056731.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J,
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
DR WPI; 99-167452/14.
DR P-PSDB; Y10885.
PT New isolated human genes encoding secreted polypeptides - useful for
PT diagnosis and treatment of pathalogical diseases
PS Claim 3; Page 269; 331pp; English.
CC The specification describes secreted proteins and their corresponding
CC polynucleotides which are useful for preventing, treating or ameliorating
CC medical conditions, e.g. by protein or gene therapy. Pathological
CC conditions can also be diagnosed by determining the amount of the
CC secreted polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the products, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, cardiovascular disorders,
CC prostate diseases, asthma, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
SQ Sequence 2351 BP; 702 A; 446 C; 518 G; 675 T;

Query Match 6.2%; Score 17; DB 1; Length 2351;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGA 274
||||||||||||||||| |
Db 2199 GGGAGGCCGAGGCAGGA 2215

RESULT 14
V11854
ID V11854 standard; DNA; 3523 BP.
AC V11854;
DT 14-SEP-1998 (first entry)
DE Human Duffy genomic DNA sequence (FY*B).
KW Duffy gp-Fy; FY*B gene; blood group; blood typing; human;
KW polymorphism; transgenic animal; hybridoma; monoclonal antibody;
KW ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT primer_bind complement (1. .23)
FT /*tag= a
FT /note= "sense primer for DNA amplification"
FT CDS 1531. .2547
FT /*tag= b
FT variation 1661
FT /*tag= c
FT /note= "G in FY*A"
FT primer_bind 3501. .3523
FT /*tag= d
FT /note= "antisense primer for DNA amplification"
PN WO9821316-A1.
PD 22-MAY-1998.
PF 14-NOV-1997; U20783.
PR 15-NOV-1996; US-749527.
PA (NYBL-) NEW YORK BLOOD CENT INC.
PI Reid ME;
DR WPI; 98-297923/26.
PT Methods of producing antibodies specific for one form of a
PT polymorphic protein - useful in blood typing etc.
PS Example 1; Fig 3A-B; 43pp; English.
CC This nucleotide sequence comprises a Duffy genomic DNA sequence
CC (FY*B) used to produce transgenic mice. It was obtained by PCR
CC amplification using FY-specific primers (see V11852-53). The
CC amplified fragment was cloned in the pBluescript vector, and a
CC purified DNA fragment containing the FY*B gene was microinjected
CC into the male pronucleus of fertilised eggs of the B6/CBA F1 mouse.
CC Transgenic mice were obtained. The invention relates to a method
CC for making monoclonal antibodies (MAbs) having pre-defined
CC specificity to an epitope characteristic of, or unique to, a single
CC form of a polymorphic protein. This includes: constructing a first
CC transgenic animal to express a first form of a polymorphic protein
CC encoded by a first allele of a gene encoding the protein;
CC constructing a second transgenic animal to express a second form of
CC the polymorphic protein encoded by a second allele of the gene
CC encoding the protein; and immunising the first transgenic animal
CC with cells from the second transgenic animal to induce an immune
CC response in the first transgenic animal yielding an antibody
CC specific for an epitope peculiar to the second form of the
CC polymorphic protein. The invention is particularly advantageous in
CC the context of making MAbs and derivative reagents specifically
CC identifying polymorphic blood group proteins, such as the Duffy
CC gp-Fy protein.
SQ Sequence 3523 BP; 720 A; 1042 C; 806 G; 955 T;

Query Match 6.2%; Score 17; DB 1; Length 3523;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 TTCTGGTCCCCACCTTT 96
||||||||||||||||| |
Db 803 TTCTGGTCCCCACCTTT 819

RESULT 15
V27017
ID V27017 standard; DNA; 3523 BP.
AC V27017;
DT 11-SEP-1998 (first entry)
DE Homo sapiens DNA fragment containing FY*B coding sequence.
KW gp-FY protein; Fyb71-81; duffy blood group; antigen; alpha; beta;
KW alternative splicing; RBC; red blood cell; malaria; treatment; ss.
OS Homo sapiens.
PN WO9821224-A1.
PD 22-MAY-1998.
PF 14-NOV-1997; U21067.
PR 15-NOV-1996; US-749543.
PA (NYBL-) NEW YORK BLOOD CENT INC.

PI Chaudhuri A, Pogo OA;
DR WPI; 98-297854/26.
PT Nucleic acid encoding gp-Fy, Duffy antigen proteins - used to
PT prevent vivax malaria and to regulate erythrocyte, neural or renal
PT function
PS Example 15; Fig 13; 87pp; English.
CC The sequence is that encoding a major subunit of the Duffy blood
CC group antigenic system, the gp-Fy proteins. The gp-Fy proteins
CC are gp-Fy alpha and gp-Fy beta which are produced from the
CC same gene via a mRNA splicing mechanism. It contains the
CC major receptor by which Plasmodium vivax enters red blood
CC cells (RBC) and causes malaria. The proteins are thus useful
CC in preventing malaria and in regulating RBC, renal and neural
CC function. The protein or certain fragments of it, may also be
CC used to generate antibodies, complementary peptides and drugs
CC modelled on their tertiary structure, useful in the same way.
SQ Sequence 3523 BP; 720 A; 1042 C; 806 G; 955 T;

ry Match 6.2%; Score 17; DB 1; Length 3523;
t Local Similarity 100.0%; Pred. No. 7.4;
tches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 TTCTGGTCCCCACCTTT 96
||||||||||||||||
Db 803 TTCTGGTCCCCACCTTT 819

Search completed: October 3, 2000, 14:37:36
Job time: 5195 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: October 3, 2000, 12:54:49 ; Search time 157.16 Seconds
(without alignments)
550.817 Million cell updates/sec

Title: US-09-065-672-4
Perfect score: 346
Sequence: 1 CTAAGGCGTGCAAACAGAGC.....CTGTCTCTATTTATACAATA 346

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

W size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	7.5	3200	1	X33947	Human HCMV inducib
c 2	24	6.9	6511	1	Q95493	Human Cdn-2 DNA. N
3	23	6.6	84	1	T25848	Human gene signatu
c 4	22	6.4	840	1	V39298	Human RAD54 nuclei
5	22	6.4	1363	1	T15455	Lung cancer specif
6	22	6.4	2310	1	Q14851	Clone pTB1283 enco
7	22	6.4	2676	1	Q14850	Clone pTB1284 enco
c 8	22	6.4	10380	1	T67164	Human alpha-N-acet
9	22	6.4	13104	1	Q46852	Clone of recombina
10	22	6.4	20303	1	T71699	Human deoxycytidyl
11	22	6.4	26764	1	T71696	Human deoxycytidyl
c 12	22	6.4	35100	1	V20441	Human c-fms oncoge
13	22	6.4	80240	1	V83940	NC-contig derived
14	22	6.4	80595	1	V83939	HC-contig derived
15	21	6.1	158	1	T25057	Human gene signatu
c 16	21	6.1	262	1	T22201	Human gene signatu
17	21	6.1	384	1	Q60667	Human brain Expres
c 18	21	6.1	423	1	Q60666	Human brain Expres
19	21	6.1	1015	1	X30159	Human secreted pro
20	21	6.1	1534	1	T18324	BRCA1 gene 5' tran
21	21	6.1	1534	1	T32611	BRCA1 gene 5' tran
22	21	6.1	3798	1	V36328	Human BRCA1 gene p
23	21	6.1	4009	1	T85827	Human interleukin-
24	21	6.1	11811	1	V83943	Bacterial artifici
25	21	6.1	24025	1	T17455	Mutated BRCA1 geno
26	21	6.1	24025	1	T17515	Mutated BRCA1 geno
27	21	6.1	24026	1	T18325	BRCA1, human breas
28	21	6.1	24026	1	T17512	Mutated BRCA1 geno
29	21	6.1	24026	1	T17513	Mutated BRCA1 geno
30	21	6.1	24026	1	T17514	Mutated BRCA1 geno
31	21	6.1	24026	1	T17516	Mutated BRCA1 geno
32	21	6.1	24026	1	T17517	Mutated BRCA1 geno
33	21	6.1	24026	1	T17518	Mutated BRCA1 geno
34	21	6.1	24026	1	T17519	Mutated BRCA1 geno
35	21	6.1	24026	1	T17521	Mutated BRCA1 geno
36	21	6.1	24026	1	T17522	Mutated BRCA1 geno
37	21	6.1	24026	1	T17523	Mutated BRCA1 geno
38	21	6.1	24026	1	T17524	Mutated BRCA1 geno
39	21	6.1	24026	1	T17526	Mutated BRCA1 geno
40	21	6.1	24026	1	T17527	Mutated BRCA1 geno
41	21	6.1	24026	1	T17528	Mutated BRCA1 geno
42	21	6.1	24026	1	T17529	Mutated BRCA1 geno
43	21	6.1	24026	1	T17530	BRCA1 genomic sequ
44	21	6.1	24026	1	T32612	BRCA1, human breas
45	21	6.1	24029	1	T17520	Mutated BRCA1 geno

ALIGNMENTS

RESULT 1
X33947
ID X33947 standard; DNA; 3200 BP.
AC X33947;
DT 30-JUN-1999 (first entry)
DE Human HCMV inducible gene, SEQ ID NO 21.
KW HCMV inducible gene; cig; human; human cytomegalovirus; interferon;
KW anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
KW drug screening; ds.
OS Homo sapiens.
PN WO9913075-A2.
PD 18-MAR-1999.
PF 08-SEP-1998; U18638.
PR 22-SEP-1997; US-059725.
PR 08-SEP-1997; US-058180.
PA (UYPR-) UNIV PRINCETON.
PI Cong J, Schenk T, Zhu H;
DR WPI; 99-243729/20.
PT New isolated human genes
PS Claim 2; Page 143-147; 184pp; English.
CC This sequence represents a human gene of the invention, that is induced
CC to express by both HCMV and interferon (IFN), designated HCMV-inducible
CC genes (cig or cigs). The invention also relates to genes that are
CC repressed in the presence of HCMV infection, designated HCMV-repressible
CC genes (crg or crgs). The products can be used to obtain agents which can
CC be used for anti-viral therapy, particularly anti-HCMV therapy. They can
CC also be used for the development of drugs that would allow for higher
CC dosage IFN treatments without the concomitant toxicity normally
CC associated with administering high levels of IFN. The products can also
CC be used for detection, diagnosis and drug screening.
SQ Sequence 3200 BP; 972 A; 629 C; 742 G; 857 T;

Query Match 7.5%; Score 26; DB 1; Length 3200;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGGGCAA 314
||||||||||||||||||||||||||
Db 380 CAGGAGTTCCAGACCAGCCTGGGCAA 405

RESULT 2
Q95493/c
ID Q95493 standard; DNA; 6511 BP.
AC Q95493;
DT 21-NOV-1995 (first entry)
DE Human Cdn-2 DNA.
KW Cdn-2; apoptosis modulator; adoptive immunotherapy; therapy; HIV;
KW autoimmune disease; reperfusion injury; hepatitis, osteoporosis;
KW shock; lymphoma; eczema; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 3312. .3947
FT /*tag= a
PN WO9515084-A.

PD 08-JUN-1995.
PF 30-NOV-1994; U13930.
PR 30-NOV-1993; US-160067.
PR 07-OCT-1994; US-320157.
PA (LXRB-) LXR BIOTECHNOLOGY INC.
PI Barr PJ, Kiefer MC;
DR WPI; 95-215106/28.
DR P-PSDB; R77877.
PT New nucleic acid sequences encoding Cdn apoptosis modulators - and
PT related vectors, transformed cells, proteins and antibodies, useful
PT or diagnosis and treatment e.g. of HIV infection, reperfusion injury
PT etc.
PS Claim 6; Fig.5A-H; 66pp; English.
CC Cdn-2 cDNA was isolated from a human placental genomic library
CC using a 950 bp fragment of Cdn-1 cDNA. Expression of Cdn-2
CC in mouse progenitor B-cell FL5.12 cells decreased IL-3-induced
CC apoptosis. The Cdn-2 protein displayed 97% amino acid identity
CC with Cdn-1 (R77876).
SQ Sequence 6511 BP; 1513 A; 1620 C; 1605 G; 1773 T;

Query Match 6.9%; Score 24; DB 1; Length 6511;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGGGC 312
||||||||||||||||||||||||
Db 1393 CAGGAGTTCCAGACCAGCCTGGGC 1370

RESULT 3
T25848
ID T25848 standard; cDNA to mRNA; 84 BP.
AC T25848;
DT 22-OCT-1996 (first entry)
DE Human gene signature HUMGS08078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1942; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 84 BP; 33 A; 17 C; 15 G; 19 T;

Query Match 6.6%; Score 23; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 0.0064;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGGG 311
|||||||||||||||||||||||
Db 14 CAGGAGTTCCAGACCAGCCTGGG 36

RESULT 4
V39298/c
ID V39298 standard; cDNA; 840 BP.
AC V39298;
DT 16-SEP-1998 (first entry)
DE Human RAD54 nucleic acid sequence comprising exon 9.
KW Human; RAD54; hRAD54; cancer; xeroderma pigmentosum; Bloom syndrome;
KW Werner's syndrome; ATR-X; diagnosis; detection; SNF2 superfamily;
KW X-linked mental retardation with alpha-thalassemia syndrome; tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN EP-844305-A2.
PD 27-MAY-1998.
PF 10-NOV-1997; 308998.
PR 13-NOV-1996; US-030676.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (UYJE-) UNIV JEFFERSON THOMAS.
PI Croce CM, Fishel RA, Rasio D, Robbins DJ;
DR WPI; 98-274189/25.
PT Human hRAD54 DNA and polypeptide - and agonists, antibodies,
PT antagonists, etc.
PS Claim 1; Page 28; 64pp; English.
CC The present sequence represents a specifically claimed partial nucleic
CC acid sequence encoding human RAD54 (hRAD54). A method for analysing a
CC sample for mutation of DNA encoding hRAD54 has been developed using a
CC DNA sequence of at least 15 and no more than 30 consecutive bases of
CC the DNA sequence encoding hRAD54. hRAD54 is a gene thought to be present
CC in tumours that display allelic imbalance at Ip32, the chromosomal band
CC identified as one of four minimal regions of chromosome 1 deletion in
CC breast carcinomas. hRAD54 is useful for production of proteins, inter
CC alia, that have been identified as novel hRAD54 by homology between the
CC amino acid sequence given in W62186 and known amino acid sequences such
CC as yeast RAD54. hRAD54 proteins are used in the treatment of cancer,
CC including Xeroderma Pigmentosum and Bloom syndrome, Werner's syndromes
CC and X-linked mental retardation with alpha-thalassaemia syndrome and
CC breast cancer. hRAD54 polynucleotides are also useful for detecting
CC complementary nucleotides for use as a diagnostic agent, especially
CC useful for diagnosis of disease or susceptibility to diseases. hRAD54
CC polynucleotide, proteins, agonists and antagonists which are proteins
CC are useful in gene therapy.
SQ Sequence 840 BP; 190 A; 200 C; 221 G; 229 T;

Query Match 6.4%; Score 22; DB 1; Length 840;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 150 CAGGAGTTCCAGACCAGCCTGG 129

RESULT 5
T15455
ID T15455 standard; DNA; 1363 BP.
AC T15455;
DT 23-APR-1996 (first entry)
DE Lung cancer specific antigen HCAVIII promoter region genomic DNA.
KW Non-small cell lung cancer; NSCLC; tumour marker; HCAVIII;
KW carbonic anhydrase; diagnosis; therapy; promoter; DNA probe;
KW fluorescent in situ hybridisation; ds.
OS Homo sapiens.
PN WO9602552-A1.
PD 01-FEB-1996.
PF 19-JUL-1995; U09145.
PR 19-JUL-1994; US-276919.

PA (CYTO-) CYTOCLONAL PHARM INC.
PI Bollon AP, Torczynski RM;
DR WPI; 96-105844/11.
PT Nucleic acid encoding the lung cancer specific antigen HCAVIII -
PT useful for diagnosis and treatment of non-small cell lung cancer
PS Claim 53; Page 62-63; 87pp; English.
CC A genomic clone (T15455) was isolated that constitutes the putative
CC promoter of the HCAVIII gene (see T15448), and probably contains
CC transcription regulatory elements directly implicated in expression
CC of HCAVIII, a cell surface antigen which is highly specific for
CC non-small cell lung carcinoma and which has features in common with
CC human carbonic anhydrases. The clone was obtd. by PCR amplification
CC using a primer pair (T15456-57) based on the putative exon 6 of the
CC HCAVIII gene. A DNA probe comprising the genomic clone plus
CC flanking sequences was used for fluorescent in situ hybridisation.
SQ Sequence 1363 BP; 352 A; 382 C; 369 G; 260 T;

Query Match 6.4%; Score 22; DB 1; Length 1363;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 554 CAGGAGTTCCAGACCAGCCTGG 575

RESULT 6
Q14851
ID Q14851 standard; DNA; 2310 BP.
AC Q14851;
DT 18-FEB-1992 (first entry)
DE Clone pTB1283 encoding complete FGF receptor.
KW Human; fibroblast growth factor; cancer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 25. .1983
FT /*tag= a
PN WO9117183-A.
PD 14-NOV-1991.
PF 25-APR-1991; J00557.
PR 27-APR-1990; JP-113146.
PR 31-JUL-1990; JP-204438.
PR 14-SEP-1990; JP-245256.
PR 28-DEC-1990; JP-415801.
PA (TAKE) TAKEDA CHEMICAL IND KK.
PI Igarashi K, Senoo M, Watanabe T;
DR WPI; 91-353723/48.
DR P-PSDB; R15269.
PT New mutein(s) of proteins - with fibroblast growth factor
PT receptor activity, useful for treating multiple endocrine
PT neoplasia, prostatic hypertrophy, used for diagnosis
PS Example 3; Fig 8; 88pp; English.
CC A cDNA library prepared from human cancer cell line Kato III mRNA
CC was screened with an oligonucleotide corresponding to amino acids
CC 529-541 of chicken basic FGF receptor. Three positive clones were
CC obtained. One was cloned into pUC118/119 to give pTB1228 (see
CC Q14848). The complete FGF coding sequence was obtained by ligating
CC the insert from pTB1228 to the DNA sequence of the plasmid pTB1281
CC insert which encodes the carboxyl terminus of the FGF receptor from
CC Glu 533 onwards.
SQ Sequence 2310 BP; 629 A; 566 C; 636 G; 479 T;

Query Match 6.4%; Score 22; DB 1; Length 2310;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 2088 CAGGAGTTCCAGACCAGCCTGG 2109

RESULT 7
Q14850
ID Q14850 standard; DNA; 2676 BP.
AC Q14850;
DT 18-FEB-1992 (first entry)
DE Clone pTB1284 encoding complete FGF receptor.
KW Human; fibroblast growth factor; cancer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 25. .2334
FT /*tag= a
PN WO9117183-A.
PD 14-NOV-1991.
PF 25-APR-1991; J00557.
PR 27-APR-1990; JP-113146.
PR 31-JUL-1990; JP-204438.
PR 14-SEP-1990; JP-245256.
PR 28-DEC-1990; JP-415801.
PA (TAKE) TAKEDA CHEMICAL IND KK.
PI Igarashi K, Senoo M, Watanabe T;
DR WPI; 91-353723/48.
DR P-PSDB; R15268.
PT New mutein(s) of proteins - with fibroblast growth factor
PT receptor activity, useful for treating multiple endocrine
PT neoplasia, prostatic hypertrophy, used for diagnosis
PS Example 3; Fig 7; 88pp; English.
CC A cDNA library prepared from human cancer cell line Kato III mRNA
CC was screened with an oligonucleotide corresponding to amino acids
CC 529-541 of chicken basic FGF receptor. Three positive clones were
CC obtained. One was cloned into pUC118/119 to give pTB1229 (see
CC Q14849). The complete FGF coding sequence was obtained by ligating
CC the insert from pTB1229 to the DNA sequence of the plasmid pTB1281
CC insert which encodes the carboxyl terminus of the FGF receptor from
CC Glu 533 onwards.
SQ Sequence 2676 BP; 743 A; 645 C; 738 G; 550 T;

Query Match 6.4%; Score 22; DB 1; Length 2676;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 2439 CAGGAGTTCCAGACCAGCCTGG 2460

RESULT 8
T67164/c
ID T67164 standard; cDNA; 10380 BP.
AC T67164;
DT 20-AUG-1997 (first entry)
DE Human alpha-N-acetylglucosaminidase gene.
KW Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;
KW gene therapy; enzyme replacement therapy; diagnosis; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 5'utr 1. .989
FT /*tag= a
FT exon 990. .1372
FT /*tag= b
FT /number= 1
FT intron 1373. .2114
FT /*tag= c
FT exon 2115. .2262
FT /*tag= d
FT /number= 2
FT intron 2263. .3055
FT /*tag= e
FT intron 3056. .3202
FT /*tag= f
FT /number= 3
FT intron 3203. .3386
FT /*tag= g

```
FT   exon            3387. .3472
FT                   /*tag= h
FT                   /number= 4
FT   intron          3473. .5666
FT                   /*tag= i
FT   exon            5667. .5923
FT                   /*tag= j
FT                   /number= 5
FT   intron          5924. .7744
FT                   /*tag= k
FT   exon            7745. .8955
FT                   /*tag= l
FT                   /number= 6
FT   3'utr           8966. .10380
FT                   /*tag= m
PN   WO9719177-A1.
PD   29-MAY-1997.
PF   22-NOV-1996; AU0747.
PR   23-NOV-1995; AU-006748.
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI   Anson DS, Blanch L, Hopwood JJ, Scott H, Weber B;
DR   WPI; 97-298114/27.
DR   P-PSDB; W18017.
PT   Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -
PT   used for the diagnosis and treatment of mucopolysaccharidosis type
PT   IIIB, also used in gene therapy
PS   Claim 8; Page 54-61; 79pp; English.
CC   A genomic DNA molecule (T67164) includes 6 exons that code for
CC   human alpha-N-acetylglucosaminidase (W18017), an enzyme that can
CC   hydrolyse the terminal alpha-N-acetylglucosamine residues at the
CC   non-reducing terminus of fragments of heparan sulphate and heparin.
CC   It was isolated by hybridisation of a human chromosome 17 library.
CC   A cDNA clone (T67163) coding for the enzyme has also been isolated.
CC   The isolated gene or cDNA, and primers/probes based on them or
CC   their complementary strands, can be used to investigate, diagnose
CC   and treat alpha-N-acetylglucosaminidase deficiency, for example in
CC   patients suffering from mucopolysaccharidosis type IIIB.
CC   Administration is by oral, i.v., i.p., enzyme replacement therapy,
CC   gene therapy or other routes.
SQ   Sequence   10380 BP;    2210 A;    2953 C;    2851 G;    2366 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 10380;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db   7439 CAGGAGTTCCAGACCAGCCTGG 7418


RESULT   9
Q46852/c
ID   Q46852 standard; DNA; 13104 BP.
AC   Q46852;
DT   26-JAN-1994  (first entry)
DE   Clone of recombinant human kappa casein gene fragment.
KW   Casein; supplement; milk; pharmaceutical; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1. .8834
FT                   /*tag= a
FT   exon            8835. .8867
FT                   /*tag= b
FT   intron          8868. .10014
FT                   /*tag= c
FT   exon            10015. .10510
FT                   /*tag= d
FT   intron          10511. .12277
FT                   /*tag= e
FT   exon            12278. .12443
FT                   /*tag= f
PN   WO9315196-A.
PD   05-AUG-1993.
PF   25-JAN-1993; DK00024.
PR   23-JAN-1992; DK-000088.
PA   (SYMB-) SYMBICOM AB.
PI   Bergstroem S, Hansson L, Hernell O, Stroemqvist M;
PI   Toernell J;
DR   WPI; 93-258675/32.
PT   DNA encoding human kappa-casein - used for obtaining recombinant
PT   polypeptide(s) for use as nutrient supplements, partic. in infant
PT   formulae
PS   Example 4; Page 92-99; 110pp; English.
CC   The recombinant human kappa casein is produced in high yields by
CC   means of either a eukaryotic or prokaryotic expression system.  It
CC   is used as a nutrient supplement in milk based products to provide a
CC   substantial improvement of the nutritional and biological value of
CC   the formulae, making it closer in similarity to human milk.  It can
CC   also be used as a pharmaceutical.
SQ   Sequence   13104 BP;    4256 A;    2497 C;    2397 G;    3953 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 13104;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db    327 CAGGAGTTCCAGACCAGCCTGG 306


RESULT  10
T71699
ID   T71699 standard; DNA; 20303 BP.
AC   T71699;
DT   20-AUG-1997  (first entry)
DE   Human deoxycytidylate deaminase intron 2 encoding DNA.
KW   Recombinant deaminase; dCMP; ds.
OS   Homo sapiens.
PN   US5622851-A.
PD   22-APR-1997.
PF   10-JAN-1995; 370975.
PR   10-JAN-1995; US-370975.
PA   (HEAL-) HEALTH RES INC.
PI   Maley F, Maley GR, Weiner KXB;
DR   WPI; 97-244391/22.
PT   DNA encoding human deoxycytidylate deaminase - for production of
PT   recombinant deaminase
PS   Claim 2; Column 83-100; 58pp; English.
CC   The present sequence encodes the human deoxycytidylate (dCMP)
CC   deaminase intron 2, which comprises 20303 base pairs from nucleotides
CC   1964-22266 of the dCMP deaminase sense strand. The dCMP deaminase gene
CC   contains a 5' untranslated region (including the promoter), 5 exons,
CC   4 introns and a 3' untranslated region (including the stop signals).
CC   The gene can be used to produce recombinant dCMP deaminase, which can
CC   be used to convert dCMP to dUMP. Also, the dCMP gene can be altered
CC   (removed or mutated) to alter DNA replication in cells, which may lead
CC   to mutagenesis.
SQ   Sequence   20303 BP;    5454 A;    4115 C;    5052 G;    5682 T;


  Query Match             6.4%;  Score 22;  DB 1;  Length 20303;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  15284 CAGGAGTTCCAGACCAGCCTGG 15305


RESULT  11
T71696
ID   T71696 standard; DNA; 26764 BP.
AC   T71696;
DT   20-AUG-1997  (first entry)
```

```
DE   Human deoxycytidylate deaminase gene.
KW   Recombinant deaminase; dCMP; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    1. .1317
FT                   /*tag= a
FT                   /note= "5' untranslated region, including promotor"
FT   exon            1318. .1425
FT                   /*tag= b
FT                   /number= 1
FT   intron          1426. .1827
FT                   /*tag= c
FT                   /number= 1
FT   exon            1828. .1963
FT                   /*tag= d
FT                   /number= 2
FT   intron          1964. .22266
FT                   /*tag= e
FT                   /number= 2
FT   exon            22267. .22383
FT                   /*tag= f
FT                   /number= 3
FT   intron          22384. .23740
FT                   /*tag= g
FT                   /number= 3
FT   exon            23741. .23837
FT                   /*tag= h
FT                   /number= 4
FT   intron          23838. .25391
FT                   /*tag= i
FT                   /number= 4
FT   exon            25392. .25467
FT                   /*tag= j
FT                   /number= 5
FT   misc_feature    25468. .26764
FT                   /*tag= k
FT                   /note= "3' untranslated region"
PN   US5622851-A.
PD   22-APR-1997.
PF   10-JAN-1995; 370975.
PR   10-JAN-1995; US-370975.
PA   (HEAL-) HEALTH RES INC.
PI   Maley F, Maley GR, Weiner KXB;
DR   WPI; 97-244391/22.
DR   P-PSDB; W18205.
PT   DNA encoding human deoxycytidylate deaminase - for production of
PT   recombinant deaminase
PS   Claim 3; Column 55-78; 58pp; English.
CC   The present sequence encodes the human deoxycytidylate (dCMP)
CC   deaminase gene, which contains a 5' untranslated region (including
CC   the promoter), 5 exons, 4 introns and a 3' untranslated region
CC   (including the stop signals). The gene can be used to produce
CC   recombinant dCMP deaminase, which can be used to convert dCMP to dUMP.
CC   Also, the dCMP gene can be altered (removed or mutated) to alter DNA
CC   replication in cells, which may lead to mutagenesis.
SQ   Sequence   26764 BP;    7079 A;    5521 C;    6539 G;    7625 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 26764;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  17247 CAGGAGTTCCAGACCAGCCTGG 17268


RESULT  12
V20441/c
ID   V20441 standard; DNA; 35100 BP.
AC   V20441;
DT   17-JUN-1998  (first entry)
DE   Human c-fms oncogene.
KW   Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
KW   cancer; antisense oligonucleotide; c-fms; ds.
OS   Homo sapiens.
PN   US5734039-A.
PD   31-MAR-1998.
PF   15-SEP-1994; 306691.
PR   15-SEP-1994; US-306691.
PA   (UYJE-) UNIV JEFFERSON THOMAS.
PI   Calabretta B, Skorski T;
DR   WPI; 98-229882/20.
PT   Anticancer composition comprising two anti-sense oligo:nucleotide(s)
PT    - targetting cytoplasmic and nuclear oncogene(s)
PS   Claim 1; Column 59-90; 92pp; English.
CC   The present sequence represents an oncogene from the present invention.
CC   The present invention describes a composition which comprises two
CC   antisense oligonucleotides. The first oligonucleotide is specific for a
CC   cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC   c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
CC   c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC   proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC   c-rel, c-vav, c-ski, c-spi, cyclin D1, PML/RAR alpha , AML1/MTG8,
CC   E2A/prl and ALL-1/AF-4. The composition is used for treating cancer.
CC   The combination of antisense oligonucleotides has synergistically
CC   enhanced ability to inhibit growth of cancer cells.
SQ   Sequence   35100 BP;    8474 A;    8597 C;    9682 G;    8347 T;


  Query Match              6.4%;  Score 22;  DB 1;  Length 35100;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    289 CAGGAGTTCCAGACCAGCCTGG 310
          ||||||||||||||||||||||
Db  33551 CAGGAGTTCCAGACCAGCCTGG 33530


RESULT  13
V83940
ID   V83940 standard; DNA; 80240 BP.
AC   V83940;
DT   03-MAR-1999  (first entry)
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS   Homo sapiens.
PN   WO9851790-A1.
PD   19-NOV-1998.
PF   13-MAY-1998; AU0352.
PR   26-AUG-1997; AU-008791.
PR   13-MAY-1997; AU-006784.
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
PI   Cancilla MR, Choo K, Du Sart D;
DR   WPI; 99-009773/01.
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy
PS   Claim 9; Fig 16A; 540pp; English.
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains
CC   an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in
CC   a compatible cell, of replicating, acting as extra-chromosomal element
CC   and segregating during cell division. The sequences can be used to
CC   construct artificial chromosomes for use in gene therapy comprising a
CC   replicable, segregating nucleic acid that confers a specific phenotype
CC   on cells. Human artificial chromosomes can propagate in human cells and
```

CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T;

Query Match 6.4%; Score 22; DB 1; Length 80240;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 27312 CAGGAGTTCCAGACCAGCCTGG 27333

RESULT 14
V[redacted]9
ID V83939 standard; DNA; 80595 BP.
AC V83939;
DT 03-MAR-1999 (first entry)
DE HC-contig derived from normal human chromosome 10q25.2 region.
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS Homo sapiens.
PN WO9851790-A1.
PD 19-NOV-1998.
PF 13-MAY-1998; AU0352.
PR 26-AUG-1997; AU-008791.
PR 13-MAY-1997; AU-006784.
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
PI Cancilla MR, Choo K, Du Sart D;
DR WPI; 99-009773/01.
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
PS Claim 8; Fig 6; 540pp; English.
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T;

Query Match 6.4%; Score 22; DB 1; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 CAGGAGTTCCAGACCAGCCTGG 310
||||||||||||||||||||||
Db 27572 CAGGAGTTCCAGACCAGCCTGG 27593

RESULT 15
T25057
ID T25057 standard; cDNA to mRNA; 158 BP.
AC T25057;
DT 11-NOV-1996 (first entry)
DE Human gene signature HUMGS07188.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1759; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 158 BP; 46 A; 35 C; 44 G; 30 T;

Query Match 6.1%; Score 21; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGGAGGCCGAGGCAGGAAGAT 278
|||||||||||||||||||||
Db 119 GGGAGGCCGAGGCAGGAAGAT 139

Search completed: October 3, 2000, 12:55:34
Job time: 7334 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: October 3, 2000, 12:57:49 ; Search time 1545.92 Seconds
(without alignments)
236.663 Million cell updates/sec

Title: US-09-065-672-3
Perfect score: 205
Sequence: 1 GCAAACAGAGCGCCACTGGG.....TACTTTGAAACATCTACTGG 205

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

size : 0

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*
44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_pr5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_vi1:*
82: gb_vi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	19	9.3	143068	11	HSU95626	U95626 Homo sapien
	2	19	9.3	216514	55	AC018744	AC018744 Oryza sat
c	3	19	9.3	225415	41	HSBA18I14	AL121928 Homo sapi
c	4	18	8.8	2698	7	YSCNACT	M23166 S.cerevisia
c	5	18	8.8	2699	5	I08122	I08122 Sequence 1
c	6	18	8.8	2724	5	I09397	I09397 Sequence 5
c	7	18	8.8	3347	7	SCNAT	X15135 Yeast NAT 1
	8	18	8.8	3530	7	SCYDL040C	Z74088 S.cerevisia
	9	18	8.8	36687	7	SCCIVL37K	Z71781 S.cerevisia
	10	18	8.8	43325	8	SPBC660	AL034563 S.pombe c
c	11	18	8.8	83536	51	AC022747	AC022747 Homo sapi
	12	18	8.8	102995	40	AL136089	AL136089 Homo sapi
c	13	18	8.8	106571	10	HS86F14	Z99572 Human DNA s
	14	18	8.8	133783	72	AC010429	AC010429 Homo sapi
c	15	18	8.8	139740	31	AP000817	AP000817 Homo sapi
	16	18	8.8	141107	67	AC022414	AC022414 Homo sapi
c	17	18	8.8	145342	69	AC023220	AC023220 Homo sapi
c	18	18	8.8	151071	31	AP001795	AP001795 Homo sapi
c	19	18	8.8	154208	78	AC021203	AC021203 Homo sapi
c	20	18	8.8	158097	54	AC008471	AC008471 Homo sapi
	21	18	8.8	159624	56	AC011021	AC011021 Homo sapi
	22	18	8.8	161624	54	AC011640	AC011640 Homo sapi
	23	18	8.8	171300	43	AC021986	AC021986 Homo sapi
	24	18	8.8	178071	67	AC024177	AC024177 Homo sapi

c	25	18	8.8	182341	32	AL139238	AL139238 Homo sapi
	26	18	8.8	182482	43	AC016703	AC016703 Homo sapi
	27	18	8.8	186243	52	AC012022	AC012022 Homo sapi
c	28	18	8.8	195832	78	AC019184	AC019184 Homo sapi
c	29	18	8.8	216215	10	HSG256O22	AL080239 Human DNA
c	30	18	8.8	240327	69	AC022422	AC022422 Homo sapi
	31	18	8.8	260270	40	AL135840	AL135840 Homo sapi
	32	17	8.3	526	13	G61963	G61963 SHGC-89514
	33	17	8.3	1155	39	AF100634	AF100634 Homo sapi
	34	17	8.3	1795	9	HSY14873	Y14873 Homo sapien
c	35	17	8.3	1850	1	ECEXPIR	X96440 E.chrysanth
	36	17	8.3	2415	9	AK001422	AK001422 Homo sapi
	37	17	8.3	2489	9	HSDARC	X85785 H.sapiens D
	38	17	8.3	2772	11	AF055992	AF055992 Homo sapi
c	39	17	8.3	2795	11	HSU43899	U43899 Human signa
	40	17	8.3	3068	10	S76830	S76830 glycoprotei
c	41	17	8.3	13271	2	AE001168	AE001168 Borrelia
	42	17	8.3	23332	42	AC014464	AC014464 Drosophil
c	43	17	8.3	23379	34	CELT08E11	AF077546 Caenorhab
	1	17	8.3	36589	9	AP001049	AP001049 Homo sapi
	3	17	8.3	39752	9	D86993	D86993 Homo sapien

ALIGNMENTS

RESULT 1
HSU95626/c
LOCUS HSU95626 143068 bp DNA PRI 16-MAY-1997
DEFINITION Homo sapiens ccr2b (ccr2), ccr2a (ccr2), ccr5 (ccr5) and ccr6 (ccr6) genes, complete cds, and lactoferrin (lactoferrin) gene, partial cds, complete sequence.
ACCESSION U95626
VERSION U95626.1 GI:2104517
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143068)
AUTHORS McCombie,W.R., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D., Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K., Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J., Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M., Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E., Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y., Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
TITLE Human BAC clone 110P12
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 143068)
AUTHORS McCombie,R.W., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D., Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K., Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J., Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M., Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E., Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y., Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1997) Advanced Genome Sequence Analysis Course, Cold Spring Harbor Laboratory, 1Bungtown Rd., Cold Spring Harbor, NY 11724, USA
COMMENT Regions with single-strand coverage are as follows:

31434 - 31443 37900 - 37968 53303 - 53357
59166 - 59206 63708 - 63998 65200 - 65335
78605 - 78713 92135 - 92137 112377 - 112551
112643 - 112778 134284 - 134309 134914 - 135019
143046 - 144068.

FEATURES Location/Qualifiers
source 1. .143068
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="BAC 110P12"
mRNA 46056. .47997
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced mRNA encoding B-form carboxyl tail. Accession Number: U80924."
/product="ccr2b"
mRNA join(46056. .47046,48255. .49505)
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced mRNA encoding A-form carboxyl tail, Accession Number U80924."
/product="ccr2a"
gene 46056. .49505
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) mRNA (Accession Number U80924), two alternatively spliced mRNAs."
CDS join(46106. .47046,48255. .48438)
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced A-form, Encoded by GenBank Accession Number U80924, gi 1168965"
/codon_start=1
/product="ccr2a"
/protein_id="AAB57791.1"
/db_xref="GI:2104518"
/translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRSLFHIALGCRIAPLQKPVCGG
PGVRPGKNVKVTTQGLLDGRGKGKSIGRAPEASLQDKEGA"
CDS 46106. .47188
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2), Accession Number: 1168965"
/codon_start=1
/product="ccr2b"
/protein_id="AAB57792.1"
/db_xref="GI:2104519"
/translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRRYLSVFFRKHITKRFCKQCPV
FYRETVDGVTSTNTPSTGEQEVSAGL"
mRNA join(59531. .59573,61472. .64785)
/gene="ccr5"
/note="confirmed by similarity to Human cc chemokine receptor 5 (ccr5) mRNA. Accession number: U54994."
/product="ccr5"
gene 59531. .64785
/gene="ccr5"
CDS 61483. .62541
/gene="ccr5"
/note="confirmed by similarity to human CC chemokine receptor 5 (ccr5) protein, encoded by GenBank Accession Number U54994, gi 1457946"
/codon_start=1
/product="ccr5"
/protein_id="AAB57793.1"
/db_xref="GI:2104520"
/translation="MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFG
FVGNMLVILILINCKRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQ
LLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFAS
LPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVICYSGIL

```
                     KTLLRCRNEKKRHRAVRLIFTIMIVYFLFWAPYNIVLLLNTFQEFFGLNNCSSSNRLD
                     QAMQVTETLGMTHCCINPIIYAFVGEKFRNYLLVFFQKHIAKRFCKCCSIFQQEAPER
                     ASSVYTRSTGEQEISVGL"
     mRNA            96634. .97683
                     /note="Putative mRNA identified by homology to CCR5 mRNA."
                     /product="ccr6"
                     /evidence=not_experimental
     gene            96642. .97676
                     /gene="ccr6"
     CDS             96642. .97676
                     /gene="ccr6"
                     /note="Identified as a gene by Grail Version 1.3c.
                     Translated sequence exhibits 42% sequence identity to CCR5
                     protein."
                     /codon_start=1
                     /evidence=not_experimental
                     /product="ccr6"
                     /protein_id="AAB57794.1"
                     /db_xref="GI:2104521"
                     /translation="MANYTLAPEDEYDVLIEGELESDEAEQCDKYDAQALSAQLVPSL
                     CSAVFVIGVLDNLLVVLILVKYKGLKRVENIYLLNLAVSNLCFLLTLPFWAHAGGDPM
                     CKILIGLYFVGLYSETFFNCLLTVQRYLVFLHKGNFFSARRRVPCGIITSVLAWVTAI
                     LATLPEYVVYKPQMEDQKYKCAFSRTPFLPADETFWKHFLTLKMNISVLVLPLFIFTF
                     LYVQMRKTLRFREQRYSLFKLVFAIMVVFLLMWAPYNIAFFLSTFKEHFSLSDCKSSY
                     NLDKSVHITKLIATTHCCINPLLYAFLDGTFSKYLCRCFHLRSNTPLQPRGQSAQGTS
                     REEPDHSTEV"
     mRNA            complement(join(124605. .124816,126528. .126717,
                     127884. .128068,130006. .130073,132023. .132164,
                     133863. .134018,135022. .135075,135890. .135980,
                     137445. .137599,138436. .138610,139077. .>139255))
                     /gene="lactoferrin"
                     /note="confirmed by similarity to lactoferrin mRNA,
                     accession number M73700"
                     /product="lactoferrin"
     gene            complement(124605. .>139255)
                     /gene="lactoferrin"
     CDS             complement(join(124782. .124816,126528. .126717,
                     127884. .128068,130006. .130073,132023. .132164,
                     133863. .134018,135022. .135075,135890. .135980,
                     137445. .137599,138436. .138610,139077. .>139253))
                     /gene="lactoferrin"
                     /note="confirmed by similarity to lactoferrin protein,
                     encoded by GenBank  Accession Number M73700, gi  186818"
                     /codon_start=1
                     /product="lactoferrin"
                     /protein_id="AAB57795.1"
                     /db_xref="GI:2104522"
                     /translation="DLSDEAERDEYELLCPDNTRKPVDKFKDCHLARVPSHAVVARSV
                     NGKEDAIWNLLRQAQEKFGKDKSPKFQLFGSPSGQKDLLFKDSAIGFSRVPPRIDSGL
                     YLGSGYFTAIQNLRKSEEEVAARRARVVWCAVGEQELRKCNQWSGLSEGSVTCSSAST
                     TEDCIALVLKGEADAMSLDGGYVYTAGKCGLVPVLAENYKSQQSSDPDPNCVDRPVEG
                     YLAVAVVRRSDTSLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQTGSCKFDEYFSQS
                     CAPGSDPRSNLCALCIGDEQGENKCVPNSNERYYGYTGAFRCLAENAGDVAFVKDVTV
                     LQNTDGNNNDAWAKDLKLADFALLCLDGKRKPVTEARSCHLAMAPNHAVVSRMDKVER
                     LKQVLLHQQAKFGRNGSDCPDKFCLFQSETKNLLFNDNTECLARLHGKTTYEKYLGPQ
                     YVAGITNLKKCSTSPLLEACEFLRK"
BASE COUNT    41194 a  30122 c  32403 g  39349 t
ORIGIN


  Query Match          9.3%;  Score 19;  DB 11;  Length 143068;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     73 CTGGTCCCCACCTTTGCAG 91
          |||||||||||||||||||
Db  32691 CTGGTCCCCACCTTTGCAG 32673


RESULT   2
AC018744
LOCUS       AC018744    216514 bp    DNA             HTG       07-MAR-2000
DEFINITION  Oryza sativa chromosome 10 clone 15022, *** SEQUENCING IN PROGRESS
            ***, 16 unordered pieces.
ACCESSION   AC018744
VERSION     AC018744.2  GI:7191023
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 216514)
  AUTHORS   McCombie,W.R.
  TITLE     Rice genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216514)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Mar 7, 2000 this sequence version replaced gi:6730690.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   158180: contig of 158180 bp in length
            *                     gap of unknown length
            *    158181   174277: contig of 16097 bp in length
            *                     gap of unknown length
            *    174278   188853: contig of 14576 bp in length
            *                     gap of unknown length
            *    188854   192653: contig of 3800 bp in length
            *                     gap of unknown length
            *    192654   196182: contig of 3529 bp in length
            *                     gap of unknown length
            *    196183   198852: contig of 2670 bp in length
            *                     gap of unknown length
            *    198853   201033: contig of 2181 bp in length
            *                     gap of unknown length
            *    201034   203123: contig of 2090 bp in length
            *                     gap of unknown length
            *    203124   205004: contig of 1881 bp in length
            *                     gap of unknown length
            *    205005   206840: contig of 1836 bp in length
            *                     gap of unknown length
            *    206841   208587: contig of 1747 bp in length
            *                     gap of unknown length
            *    208588   210229: contig of 1642 bp in length
            *                     gap of unknown length
            *    210230   211854: contig of 1625 bp in length
            *                     gap of unknown length
            *    211855   213466: contig of 1612 bp in length
            *                     gap of unknown length
            *    213467   215011: contig of 1545 bp in length
            *                     gap of unknown length
            *    215012   216514: contig of 1503 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216514
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /clone="15022"
BASE COUNT    62170 a  45887 c  47331 g  60900 t    226 others
ORIGIN


  Query Match          9.3%;  Score 19;  DB 55;  Length 216514;
  Best Local Similarity   100.0%;  Pred. No. 4.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    166 CTCTCTAATAAGAAAACAT 184
          ||||||||||||||||||
Db 137518 CTCTCTAATAAGAAAACAT 137536


RESULT   3
HSBA18I14/c
LOCUS       HSBA18I14  225415 bp    DNA             HTG       20-APR-2000
DEFINITION  Homo sapiens chromosome 10 clone RP11-18I14, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL121928
VERSION     AL121928.8  GI:7635624
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 225415)
  AUTHORS   Wilson,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 22, 2000 this sequence version replaced gi:7452949.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA18I14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 15% of reads
            Sequencing vector: plasmid; L08752; 84% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 203428 bases at least Q40
            Consensus quality: 210595 bases at least Q30
            Consensus quality: 215972 bases at least Q20
            Insert size: 221115; sum-of-contigs
            Insert size: 164357; 5.4% error; agarose-fp
            Quality coverage: 6.37x in Q20 bases; sum-of-contigs Quality
            coverage: 8.57x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently * consists
            of 44 contigs. The true order of the pieces is * not known and
            their order in this sequence record is * arbitrary.  Where the
            contigs adjacent to the vector can * be identified, they are
            labelled with 'clone_end' in the * feature table.  Some order and
            orientation information * can tentatively be deduced from paired
            sequencing reads * which have been identified to span the gap
            between two * contigs.  These are labelled as part of the same *
            'fragment_chain', and the order and relative orientation * of the
            pieces within a fragment_chain is reflected in * this file.  Gaps
            between the contigs are represented as * runs of N, but the exact
            sizes of the gaps are unknown. * This record will be updated with
            the finished sequence as * soon as it is available and the
            accession number will be * preserved.
            *       1     7454 contig of 7454 bp in length; fragment_chain 1 *
                7555    33234 contig of 25680 bp in length; fragment_chain 1 *
               33335    68872 contig of 35538 bp in length; fragment_chain 1 *
               68973    71980 contig of 3008 bp in length; fragment_chain 2 *
               72081    73191 contig of 1111 bp in length; fragment_chain 2 *
               73292    75668 contig of 2377 bp in length; fragment_chain 2 *
               75769   106651 contig of 30883 bp in length; fragment_chain 3 *
              106752   142041 contig of 35290 bp in length; fragment_chain 3 *
              142142   153747 contig of 11606 bp in length; fragment_chain 3 *
              153848   163721 contig of 9874 bp in length; fragment_chain 4 *
              163822   175370 contig of 11549 bp in length; fragment_chain 4 *
              175471   176510 contig of 1040 bp in length; fragment_chain 5 *
              176611   177620 contig of 1010 bp in length; fragment_chain 5 *
              177721   179026 contig of 1306 bp in length; fragment_chain 6 *
              179127   181522 contig of 2396 bp in length; fragment_chain 6 *
              181623   182876 contig of 1254 bp in length; fragment_chain 7 *
              182977   184902 contig of 1926 bp in length; fragment_chain 7 *
              185003   186125 contig of 1123 bp in length
              *   186226   187353 contig of 1128 bp in length
              *   187454   188976 contig of 1523 bp in length
              *   189077   190281 contig of 1205 bp in length
              *   190382   191893 contig of 1512 bp in length
              *   191994   193218 contig of 1225 bp in length
              *   193319   195372 contig of 2054 bp in length
              *   195473   196928 contig of 1456 bp in length
              *   197029   198538 contig of 1510 bp in length
              *   198639   199815 contig of 1177 bp in length
              *   199916   201566 contig of 1651 bp in length
              *   201667   202994 contig of 1328 bp in length
              *   203095   204580 contig of 1486 bp in length
              *   204681   206583 contig of 1903 bp in length
              *   206684   208153 contig of 1470 bp in length
              *   208254   209357 contig of 1104 bp in length
              *   209458   210641 contig of 1184 bp in length
              *   210742   211804 contig of 1063 bp in length
              *   211905   213296 contig of 1392 bp in length
              *   213397   214984 contig of 1588 bp in length
              *   215085   216477 contig of 1393 bp in length
              *   216578   217599 contig of 1022 bp in length
              *   217700   218732 contig of 1033 bp in length
              *   218833   220111 contig of 1279 bp in length
              *   220212   221249 contig of 1038 bp in length
              *   221350   223431 contig of 2082 bp in length
              *   223532   225415 contig of 1884 bp in length.
              * NOTE: This is a 'working draft' sequence.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
FEATURES             Location/Qualifiers
     source          1. .225415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-18I14"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .7454
                     /note="assembly_fragment:03946
                     fragment_chain:1"
     misc_feature    7555. .33234
                     /note="assembly_fragment:04346
                     fragment_chain:1"
     misc_feature    33335. .68872
                     /note="assembly_fragment:00300
                     fragment_chain:1"
     misc_feature    68973. .71980
                     /note="assembly_fragment:04457
                     fragment_chain:2"
     misc_feature    72081. .73191
                     /note="assembly_fragment:04755
                     fragment_chain:2"
     misc_feature    73292. .75668
                     /note="assembly_fragment:05108
                     fragment_chain:2"
     misc_feature    75769. .106651
                     /note="assembly_fragment:04582
                     fragment_chain:3"
     misc_feature    106752. .142041
                     /note="assembly_fragment:02140
                     fragment_chain:3"
     misc_feature    142142. .153747
                     /note="assembly_fragment:04188
                     fragment_chain:3"
     misc_feature    153848. .163721
                     /note="assembly_fragment:00668
                     fragment_chain:4"
     misc_feature    163822. .175370
                     /note="assembly_fragment:03147
                     fragment_chain:4"
```

```
     misc_feature    175471. .176510
                     /note="assembly_fragment:04690
                     fragment_chain:5"
     misc_feature    176611. .177620
                     /note="assembly_fragment:04994
                     fragment_chain:5"
     misc_feature    177721. .179026
                     /note="assembly_fragment:04809
                     fragment_chain:6"
     misc_feature    179127. .181522
                     /note="assembly_fragment:04936
                     fragment_chain:6"
     misc_feature    181623. .182876
                     /note="assembly_fragment:04916
                     fragment_chain:7"
     misc_feature    182977. .184902
                     /note="assembly_fragment:05043
                     fragment_chain:7"
     misc_feature    185003. .186125
                     /note="assembly_fragment:00253"
     misc_feature    186226. .187353
                     /note="assembly_fragment:00391"
     misc_feature    187454. .188976
                     /note="assembly_fragment:00453"
     misc_feature    189077. .190281
                     /note="assembly_fragment:00825"
     misc_feature    190382. .191893
                     /note="assembly_fragment:00938"
     misc_feature    191994. .193218
                     /note="assembly_fragment:01600"
     misc_feature    193319. .195372
                     /note="assembly_fragment:02835"
     misc_feature    195473. .196928
                     /note="assembly_fragment:03033"
     misc_feature    197029. .198538
                     /note="assembly_fragment:03415"
     misc_feature    198639. .199815
                     /note="assembly_fragment:03422"
     misc_feature    199916. .201566
                     /note="assembly_fragment:04573"
     misc_feature    201667. .202994
                     /note="assembly_fragment:04611"
     misc_feature    203095. .204580
                     /note="assembly_fragment:04616"
     misc_feature    204681. .206583
                     /note="assembly_fragment:04623"
     misc_feature    206684. .208153
                     /note="assembly_fragment:04624"
     misc_feature    208254. .209357
                     /note="assembly_fragment:04631"
     misc_feature    209458. .210641
                     /note="assembly_fragment:04633"
     misc_feature    210742. .211804
                     /note="assembly_fragment:04634"
     misc_feature    211905. .213296
                     /note="assembly_fragment:04644"
     misc_feature    213397. .214984
                     /note="assembly_fragment:04694"
     misc_feature    215085. .216477
                     /note="assembly_fragment:04743"
     misc_feature    216578. .217599

  Query Match             9.3%;  Score 19;  DB 41;  Length 225415;
  Best Local Similarity   100.0%;  Pred. No. 4.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      35 AGGCCGATGCTTGCTTGCA 53
           |||||||||||||||||||
Db   85879 AGGCCGATGCTTGCTTGCA 85861


RESULT   4
YSCNACT/c
```

```
LOCUS       YSCNACT      2698 bp    mRNA            PLN       16-FEB-1996
DEFINITION  S.cerevisiae N-acetyltransferase (AAA1) mRNA, complete cds.
ACCESSION   M23166 J04837
VERSION     M23166.1  GI:172027
KEYWORDS    N-acetyltransferase.
SOURCE      Saccharomyces cerevisiae (strain TD71.8) (clone: pBN9) cDNA to
            mRNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2698)
  AUTHORS   Lee,F.J., Lin,L.W. and Smith,J.A.
  TITLE     Molecular cloning and sequencing of a cDNA encoding N
            alpha-acetyltransferase from Saccharomyces cerevisiae
  JOURNAL   J. Biol. Chem. 264 (21), 12339-12343 (1989)
  MEDLINE   89308659
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            F.-J.Lee, 10-APR-1989.
FEATURES             Location/Qualifiers
     source          1. .2698
                     /organism="Saccharomyces cerevisiae"
                     /strain="TD71.8"
                     /db_xref="taxon:4932"
                     /clone="pBN9"
                     /note="(vector lambda gt11)"
     gene            1. .2698
                     /gene="NAT1"
                     /db_xref="SGD:S0002198"
     mRNA            <22. .2698
                     /gene="NAT1"
                     /note="N-acetyltransferase mRNA"
                     /db_xref="SGD:S0002198"
     CDS             22. .2586
                     /gene="NAT1"
                     /note="N-acetyltransferase"
                     /codon_start=1
                     /db_xref="SGD:S0002198"
                     /protein_id="AAA88728.1"
                     /db_xref="GI:172028"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
BASE COUNT      926 a    491 c    533 g    748 t
ORIGIN      Chromosome 4.


  Query Match             8.8%;  Score 18;  DB 7;  Length 2698;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 TGCTTTTCTCTCTAATAA 176
           ||||||||||||||||||
Db     765 TGCTTTTCTCTCTAATAA 748


RESULT   5
I08122/c
LOCUS       I08122       2699 bp                    PAT       02-DEC-1994
DEFINITION  Sequence 1 from Patent EP 0334004.
ACCESSION   I08122
VERSION     I08122.1  GI:589163
```

```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2699)
  AUTHORS   Smith,J.A. and Lee,F-J.S.
  TITLE     Isolation, purification, characterization, cloning and sequencing
            of N alpha-acetyltransferase
  JOURNAL   Patent: EP 0334004-A1 1 27-SEP-1989;
FEATURES             Location/Qualifiers
     source          1. .2699
                     /organism="unknown"
BASE COUNT      927 a    492 c    532 g    748 t
ORIGIN


  Query Match             8.8%;  Score 18;  DB 5;  Length 2699;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Q      159 TGCTTTTCTCTCTAATAA 176
           ||||||||||||||||||
Db     765 TGCTTTTCTCTCTAATAA 748


RESULT   6
I09397/c
LOCUS       I09397       2724 bp                    PAT       02-DEC-1994
DEFINITION  Sequence 5 from Patent WO 8907138.
ACCESSION   I09397
VERSION     I09397.1  GI:587894
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2724)
  AUTHORS   Smith,J.A. and Lee,F-J.S.
  JOURNAL   Patent: WO 8907138-A 5 10-AUG-1989;
FEATURES             Location/Qualifiers
     source          1. .2724
                     /organism="unknown"
BASE COUNT      952 a    491 c    533 g    748 t
ORIGIN


  Query Match             8.8%;  Score 18;  DB 5;  Length 2724;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 TGCTTTTCTCTCTAATAA 176
           ||||||||||||||||||
Db     765 TGCTTTTCTCTCTAATAA 748


RESULT   7
SCNAT/c
LOCUS       SCNAT        3347 bp    DNA             PLN       12-SEP-1993
DEFINITION  Yeast NAT 1 gene for N-terminal acetyltransferase.
ACCESSION   X15135
VERSION     X15135.1  GI:4027
KEYWORDS    acetyltransferase; NAT 1 gene.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3347)
  AUTHORS   Grunstein,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1989) Grunstein M., UCLA, Biology Department, Los
            Angeles CA 90024, USA
REFERENCE   2  (bases 1 to 3347)
  AUTHORS   Mullen,J.R., Kayne,P.S., Moerschell,R.P., Tsunasawa,S.,
            Gribskov,M., Colavito-Shepanski,M., Grunstein,M., Sherman,F. and
            Sternglanz,R.
  TITLE     Identification and characterization of genes and mutants for an
            N-terminal acetyltransferase from yeast
  JOURNAL   EMBO J. 8 (7), 2067-2075 (1989)
  MEDLINE   90005412
COMMENT     See <X01419> for overlapping sequence.
FEATURES             Location/Qualifiers
     source          1. .3347
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
     CDS             337. .2901
                     /note="acetyltransferase (AA 1-854)"
                     /codon_start=1
                     /protein_id="CAA33233.1"
                     /db_xref="GI:4028"
                     /db_xref="SWISS-PROT:P12945"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
     CDS             3330. .>3347
                     /note="SIR2 (AA 1-6)"
                     /codon_start=1
                     /protein_id="CAA33234.1"
                     /db_xref="GI:578200"
                     /translation="MTIPHM"
BASE COUNT     1127 a    625 c    653 g    942 t
ORIGIN


  Query Match             8.8%;  Score 18;  DB 7;  Length 3347;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 TGCTTTTCTCTCTAATAA 176
           ||||||||||||||||||
Db    1080 TGCTTTTCTCTCTAATAA 1063


RESULT   8
SCYDL040C
LOCUS       SCYDL040C    3530 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome IV reading frame ORF YDL040c.
ACCESSION   Z74088 Z71256
VERSION     Z74088.1  GI:1431024
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3530)
  AUTHORS   Paulin,L., Saren,A.M. and Laamanen,P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3530)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1996) Data collected by MIPS on behalf of the
            European yeast chromosome IV sequencing project. MIPS, at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .3530
```

```
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
     gene            complement(419. .2983)
                     /gene="NAT1"
                     /db_xref="SGD:S0002198"
     CDS             complement(419. .2983)
                     /gene="NAT1"
                     /note="ORF YDL040c"
                     /codon_start=1
                     /db_xref="SGD:S0002198"
                     /protein_id="CAA98599.1"
                     /db_xref="GI:1431025"
                     /db_xref="SWISS-PROT:P12945"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
BASE COUNT     1009 a    672 c    660 g    1189 t
ORIGIN


  Query Match             8.8%;  Score 18;  DB 7;  Length 3530;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    159 TGCTTTTCTCTCTAATAA 176
          ||||||||||||||||||
Db   2240 TGCTTTTCTCTCTAATAA 2257


RESULT   9
SCCIVL37K
LOCUS       SCCIVL37K   36687 bp    DNA             PLN       14-FEB-1997
DEFINITION  S.cerevisiae chromosome IV left arm (EU) DNA segment (36687 bp).
ACCESSION   Z71781
VERSION     Z71781.1  GI:1279667
KEYWORDS    act2 gene; actin; FAD synthetase; FAD1 gene; MPS1 gene; NAM1 gene;
            NAT 1 gene; protein kinase; protein phosphatase; SIR2 gene; SIT4
            gene.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 36687)
  AUTHORS   Saren,A.M., Laamanen,P., Lejarcegui,J.B. and Paulin,L.
  TITLE     The sequence of a 36.7 kb segment on the left arm of chromosome IV
            from Saccharomyces cerevisiae reveals 20 non-overlapping open
            reading frames (ORFs) including SIT4, FAD1, NAM1, RNA11, SIR2,
            NAT1, PRP9, ACT2 and MPS1 and 11 new ORFs
  JOURNAL   Yeast 13 (1), 65-71 (1997)
  MEDLINE   97197972
REFERENCE   2  (bases 1 to 36687)
  AUTHORS   Paulin,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1996) Paulin L., Institute of Biotechnology, DNA
            sequencing & Synthesis Laboratory, Biocentre 1, P.O.BOX 56
            (Viikinkaari 9), FIN-00014 University of Helsinki, Finland
FEATURES             Location/Qualifiers
     source          1. .36687
                     /organism="Saccharomyces cerevisiae"
                     /strain="alpha S288c"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
                     /clone="cosmid 2G24"
                     /clone_lib="chromosome library of C. Jacq (Paris)"
     CDS             complement(<1. .556)
                     /standard_name="D2690"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96441.1"
                     /db_xref="GI:1279668"
                     /db_xref="SPTREMBL:Q07351"
                     /translation="MLVSSSFASSIDSVMMLVSSSFASSIDSVMSHETMSLRRNPPFI
                     DTPEKMPNPTASPNGTIHHLIDPSLPLLSSTTSSSRSTLSSTLNSPPPPPLTTSYSSY
                     NSSACQSITSSPTDNTALAHNSKCYFPHSLSPTPLSSNSSSHVILPPISSFTNLITVA
                     EREFNGRSNSLHANFTSPVPRTVLD"
     gene            2114. .3049
                     /gene="SIT4"
                     /db_xref="SGD:S0002205"
     CDS             2114. .3049
                     /gene="SIT4"
                     /standard_name="D2693"
                     /note="Acc.no. M24395"
                     /codon_start=1
                     /db_xref="SGD:S0002205"
                     /product="protein phosphatase catalytic subunit homologue
                     SIT4"
                     /protein_id="CAA96442.1"
                     /db_xref="GI:1279669"
                     /db_xref="SWISS-PROT:P20604"
                     /translation="MVSRGPDEWLETIKKCQALTENEMKQLCEMVKELLMEESNIQPV
                     QTPVTVCGDIHGQFHDLLELFRTAGGFPDDINYIFLGDYVDRGYYSLETFTLLMCLKV
                     KYPAKITLVRGNHESRQITQVYGFYEECLNKYGSTTVWKYCCQVFDFLTLAAIIDGKI
                     LCVHGGLSPEIRMLDQIRVLSRAQEVPHEGGFSDLLWSDPDNVEAWQVSPRGAGWLFG
                     SKVAREFNHVNGLNLIARAHQLVMEGFKYHFPEKDVVTVWSAPNYCYRCGNVASVMKV
                     DEDLEPTFKIFSAVPDDYIRESTANHNNQRAGYFL"
     CDS             3582. .4103
                     /standard_name="D2699"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96443.1"
                     /db_xref="GI:1279670"
                     /db_xref="SPTREMBL:Q12408"
                     /translation="MTHSLKALFALLFLYTAAVNAGVIGIFNALPPPNTKPINGESPL
                     YQCDILDKQLVEIKEVNLDPNPPVRGENLTISANGEVFETIEEGAYIDVEVRLGYIRL
                     LSQTFDLCETLEDNDIEGLSCPIEPGEYNIKKIVEIPGEVPPGKYVVVARAYTEKDDL
                     ITCLTGEVIFPPR"
     gene            complement(5030. .5950)
                     /gene="FAD1"
                     /db_xref="SGD:S0002203"
     CDS             complement(5030. .5950)
                     /gene="FAD1"
                     /standard_name="D2702"
                     /note="Acc.no. U12331"
                     /codon_start=1
                     /db_xref="SGD:S0002203"
                     /product="FAD synthetase"
                     /protein_id="CAA96444.1"
                     /db_xref="GI:1279671"
                     /db_xref="SWISS-PROT:P38913"
                     /translation="MQLSKAAEMCYEITNSYLHIDQKSQIIASTQEAIRLTRKYLLSE
                     IFVRWSPLNGEISFSYNGGKDCQVLLLLYLSCLWEYFFIKAQNSQFDFEFQSFPMQRL
                     PTVFIDQEETFPTLENFVLETSERYCLSLYESQRQSGASVNMADAFRDFIKIYPETEA
                     IVIGIRHTDPFGEALKPIQRTDSNWPDFMRLQPLLHWDLTNIWSFLLYSNEPICGLYG
                     KGFTSIGGINNSLPNPHLRKDSNNPALHFEWEIIHAFGKDAEGERSSAINTSPISVVD
                     KERFSKYHDNYYPGWYLVDDTLERAGRIKN"
     gene            complement(6581. .7630)
                     /gene="NAM1"
                     /db_xref="SGD:S0002202"
     CDS             complement(6581. .7630)
                     /gene="NAM1"
                     /standard_name="D2705"
                     /codon_start=1
                     /db_xref="SGD:S0002202"
```

```
                     /protein_id="CAA96445.1"
                     /db_xref="GI:1279672"
                     /db_xref="SWISS-PROT:P10849"
                     /translation="MIRTSSILKNCNYRYIHCIHRCLLNEANLKDRKTHNVERVSNEK
                     TFEQALEEERKVFGELFEAGARVENMRHTNASKIIDKYYNGLQDNSEGTSVKKEKIVF
                     NHSQRAQRKLPNKDHEFLKETAGNDYVYERAEPSAISTKTISEQTRTLLEKIFDEDNS
                     INKSNRELLNLNLRKGSGMEALRQPVAHSNVKFSEEVMQEIGNKIRYQTTLDQVLEPH
                     IDYLREAVKSDYDLLRYLKQSLDIYKKRNKDLELKMNAESSNIFEDIRSACINKPAEL
                     PKPLAMTLPYIIVKSLRLGDFDFPADRKYTLISYVYNECKNNMDASLYLTICNVDFYN
                     LLVQLLWETSRRFAT"
gene                 complement(8021. .8812)
                     /gene="RNA11"
                     /db_xref="SGD:S0002201"
CDS                  complement(8021. .8812)
                     /gene="RNA11"
                     /standard_name="D2711"
                     /note="Protein sequence is in conflict with the conceptual
                     translation; Acc.no. X01419"
                     /codon_start=1
                     /db_xref="SGD:S0002201"
                     /product="RNA11"
                     /protein_id="CAA96446.1"
                     /db_xref="GI:1279673"
                     /db_xref="SWISS-PROT:Q07350"
                     /translation="LEGVGSKKGGGGIASESQFNLQRRKEVESLLSKGENVPYTFQDE
                     KDDQVRSNPYIYKNHSGKLVCKLCNTMHMSWSSVERHLGGKKHGLNVLRRGISIEKSS
                     LGREGQTTHDFRQQQKIIEAKQSLKNNGTIPVCKIATVKNPKNGSVGLAIQVNYSSEV
                     KENSVDSDDKAKVPPLIRIVSGLELSDTKQKGKKFLVIAYEPFENIAIELPPNEILFS
                     ENNDMDNNNDGVDELNKKCTFWDAISKLYYVQFFFKQAEQEQADV"
gene                 complement(9098. .10786)
                     /gene="SIR2"
                     /db_xref="SGD:S0002200"
CDS                  complement(9098. .10786)
                     /gene="SIR2"
                     /standard_name="D2714"
                     /note="Acc.no. X01419"
                     /codon_start=1
                     /db_xref="SGD:S0002200"
                     /product="SIR2"
                     /protein_id="CAA96447.1"
                     /db_xref="GI:1279674"
                     /db_xref="SWISS-PROT:P06700"
                     /translation="MTIPHMKYAVSKTSENKVSNTVSPTQDKDAIRKQPDDIINNDEP
                     SHKKIKVAQPDSLRETNTTDPLGHTKAALGEVASMELKPTNDMDPLAVSAASVVSMSN
                     DVLKPETPKGPIIISKNPSNGIFYGPSFTKRESLNARMFLKYYGAHKFLDTYLPEDLN
                     SLYIYYLIKLLGFEVKDQALIGTINSIVHINSQERVQDLGSAISVTNVEDPLAKKQTV
                     RLIKDLQRAINKVLCTRLRLSNFFTIDHFIQKLHTARKILVLTGAGVSTSLGIPDFRS
                     SEGFYSKIKHLGLDDPQDVFNYNIFMHDPSVFYNIANMVLPPEKIYSPLHSFIKMLQM
                     KGKLLRNYTQNIDNLESYAGISTDKLVQCHGSFATATCVTCHWNLPGERIFNKIRNLE
                     LPLCPYCYKKRREYFPEGYNNKVGVAASQGSMSERPPYILNSYGVLKPDITFFGEALP
                     NKFHKSIREDILECDLLICIGTSLKVAPVSEIVNMVPSHVPQVLINRDPVKHAEFDLS
                     LLGYCDDIAAMVAQKCGWTIPHKKWNDLKNKNFKCQEKDKGVYVVTSDEHPKTL"
CDS                  10443. .10796
                     /standard_name="D2717"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96448.1"
                     /db_xref="GI:1279675"
                     /db_xref="SPTREMBL:Q12352"
                     /translation="MIGPLGVSGFKTSLDIDTTEAADTAKGSMSLVGLSSIDATSPSA
                     ALVCPSGSVVLVSLKESGCATFIFLCEGSSLFIMSSGCFLIASLSCVGLTVFETLFSL
                     VFDTAYFICGMVIQL"
gene                 complement(11215. .13779)
                     /gene="NAT1"
                     /db_xref="SGD:S0002198"
CDS                  complement(11215. .13779)
                     /gene="NAT1"
                     /standard_name="D2720"
                     /note="Acc.no. X15135"
                     /codon_start=1
                     /db_xref="SGD:S0002198"
                     /product="N-terminal acetyltransferase"
                     /protein_id="CAA96449.1"
                     /db_xref="GI:1279676"
                     /db_xref="SWISS-PROT:P12945"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
CDS                  complement(14327. .14674)
                     /standard_name="D2723"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96450.1"
                     /db_xref="GI:1279677"
                     /db_xref="SPTREMBL:Q12459"
                     /translation="MTSPASTSTISNVQSTASVMNHSIEDNISAAASLESVSGTSTKD
```

Query Match 8.8%; Score 18; DB 7; Length 36687;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    159 TGCTTTTCTCTCTAATAA 176
          ||||||||||||||||||
Db  13036 TGCTTTTCTCTCTAATAA 13053
```

```
RESULT  10
SPBC660
LOCUS       SPBC660     43325 bp    DNA             PLN       31-JAN-2000
DEFINITION  S.pombe chromosome II cosmid c660.
ACCESSION   AL034563
VERSION     AL034563.1  GI:4049499
KEYWORDS    6-phosphogluconate dehydrogenase decarboxylating; cell wall
            protein; class v pyridoxal phosphate dependent aminotransferase;
            elongation factor g; elongation factor Tu family; fbp1;
            fructose-1,6-bisphosphatase; G beta repeat; glycine-rich protein;
            low-complexity gene-free region; mik1; mitosis inhibitor protein
            kinase mik1; myb like dna-binding protein; neutral trehalase; ntp1;
            polyA-binding protein; ras11; replication factor-a protein 1;
            ribonucleoprotein; RNA recognition; RNA3' Cleavage factor Ib; rpa1;
            ssb1; transcription initiation factor iif beta subunit; WD domain;
            yeast CF Ib.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
            Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE   1  (bases 1 to 43325)
  AUTHORS   Lyne,M., Rajandream,M.A., Barrell,B.G. and Rieger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-1998) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
            and Biotechnologische und molekularbiologische Forschung,
            Angelhofweg 39, D-69259 Wilhelmsfeld, Germany
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            During 1995 to 1996 about 66% of S. pombe chromosome 1 was
            sequenced by the Sanger Centre.  The sequencing of the S. pombe
            genome is now being continued with funding from The European
            Commission. Fourteen European sequencing laboratories, including
            the Sanger Centre, are participating in the project.
            Protein coding regions (CDS) have been predicted with the help of
```

computer analysis using the Genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid c660 is overlapped at the 3' end by cosmid 1198 (contained in EMBL entry SP33010 accession number U33010).

FEATURES Location/Qualifiers

source 1. .43325
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="II"
/clone="cosmid c660"
/map="IIL"

gene complement(1. .1482)
/gene="SPBC660.01c"

CDS complement(join(1. .86,138. .464,506. .1482))
/partial
/gene="SPBC660.01c"
/note="SPBC660.01c, SIMILARITY:Schizosaccharomyces pombe, CAB52717, putative myb-like dna-binding protein., (496 aa), fasta scores: opt: 478, E():4.7e-23, (30.6% identity in 350 aa)"
/codon_start=1
/label=SPBC660.01c
/product="putative myb like dna-binding protein"
/protein_id="CAA22521.1"
/db_xref="GI:4049500"
/db_xref="SPTREMBL:O94422"
/translation="MDTSVLNPELQIHGFIGVDSLQSSRKRKNDFDDFPLNKGLKTNN
NDYSGSIEPKFSPALSIKEDGKNDRNFEALMSLQAQDSNLSEQNTSIHLDALASSSIA
LGNDNVDSSAFLSKVNKGVNAMRNSTSNQTNDSILISPSEITNMDPFLKGSARWTAEH
WDYLERRMQNFCQTYSLDHTQVADSLHEKRLHGPLSSLVKLLVQEMPSFTRRTILRHL
RALYNIPGYEKYSRKNSSGRGDFGVQETAIISQEVHNFIMDQGWSEYQFCNQIWAGKC
PKTIRMFYSNLYKKLSHRDAKSIYHHVRRAYNPFEDRCVWSKEEDEELRKNVVEHGKC
WTKIGRKMARMPNDCRDRWRDVVRFGDKLKRNAWSLEEETQLLQIVAELRNREDLSSD
INWTLVAQMLGTRTRLQCRYKFQQLTKAASKFELQENVWLLERIYDSLLNNGGKIHWE
NIVKEANGRWTRD"

misc_feature 1. .7073
/note="nominal overlap with cosmids SP33010, EM:U33010 S. pombe chromosome 2"

misc_feature complement(87. .104)
/gene="SPBC660.01c"
/note="ctaatattttaatttaag, splice branch and acceptor"

misc_feature complement(132. .137)
/gene="SPBC660.01c"
/note="gtaagt, splice donor sequence"

misc_feature complement(join(383. .464,506. .537))
/gene="SPBC660.01c"
/note="Match to PF00249 myb_DNA-binding, Myb-like DNA-binding domain Score 30.86"

misc_feature complement(465. .475)
/gene="SPBC660.01c"
/note="ttaacgtttag, splice branch and acceptor"

misc_feature complement(500. .505)
/gene="SPBC660.01c"
/note="gtatgt, splice donor sequence"

gene 2618. .3929
/gene="SPBC660.02"

CDS join(2618. .2738,2785. .3929)
/gene="SPBC660.02"
/note="SPBC660.02, len:421, SIMILARITY:Schizosaccharomyces pombe, O13286, srw1., (556 aa), fasta scores: opt: 1364, E():0, (50.6% identity in 385 aa)"
/codon_start=1
/label=SPBC660.02
/product="WD domain; G beta repeat protein"
/protein_id="CAA22522.1"
/db_xref="GI:4049501"
/db_xref="SPTREMBL:O94423"
/translation="MGDRFIPIRNVSNEFNFSFQSFKECVLSHGSNLRRKTSGTIQRQ
FMELLSMELFGSQASRSRAFYYGEDKRKIEKKMLDTPDRKSYSLSPISPQSQDMLRQP
QKPKRAFPKTPYKILDAPYLKNDFYLNLLDWGQSNVLAVGLASSIYLWSAASGKVVQL
HDFGATNHVTSVLWTGKGTQLAVGTDSGVIYIWDIESTKSVRSLKGHSERVAALAWND
NTLTSGGKDEVILHHDLRAPGCCAEMMKVHEQEICGLQWDRSLGQLASGGNDNNLFVW
DYRSSRPLHKFEEHTAAVKAIGWSPHQRGILASGGGTIDRCLTIHNTLTGRLQNKLDT
GSQVCNMAWSKTSNEIVTTHGFAKNQVSLWKYPSLKNIANLTAHTNRVLYLSMSPDGQ
SIVTGAGDETLRFWKLFNKKPKEESTLIR"

misc_feature 2739. .2744
/gene="SPBC660.02"
/note="gtatgt, splice donor sequence"

misc_feature 2772. .2784
/gene="SPBC660.02"
/note="ctaacgacagcag, splice branch and acceptor"

misc_feature 3258. .3353
/gene="SPBC660.02"
/note="Match to PF00400 WD40, WD domain, G-beta repeat Score 22.93"

misc_feature 3378. .3494
/gene="SPBC660.02"
/note="Match to PF00400 WD40, WD domain, G-beta repeat Score 31.68"

misc_feature 3771. .3884
/gene="SPBC660.02"
/note="Match to PF00400 WD40, WD domain, G-beta repeat Score 20.22"

gene complement(3930. .5015)
/gene="SPBC660.03c"

CDS complement(join(3930. .4184,4288. .4431,4491. .5015))
/gene="SPBC660.03c"
/note="SPBC660.03c, len:307, SIMILARITY:Saccharomyces cerev isiae, YGR005C, T2FB_YEAST, transcription initiation factor iif, beta subunit, (400 aa), fasta scores: opt: 461, E():2 .9e-32, (31.0% identity in 368 aa); SPBC660.03c, len:307, SIMILARITY:Saccharomyces cerevisiae, T2FB_YEAST, transcription initiation factor iif, beta subunit, (400 aa), fasta scores: opt: 461, E():1.4e-22, (31.0% identity in 368 aa)"
/codon_start=1
/label=SPBC660.03c
/product="transcription initiation factor iif, beta subunit"
/protein_id="CAA22523.1"
/db_xref="GI:4049502"
/db_xref="SPTREMBL:O94424"
/translation="MSEEKPTVRTEEDDRYEDDAGDLDLGQIGSRVWLVKIPKFLMDK
WNSIPEDDAANLGCVRVKNDEIQLLLQNSPENADVPKIYNLRVMNKFVRNSYVFRESE
TSSSMKSTALVGTVAHECNVSPVINDDYRRVMQKRALAASAPKRKVQMIDDRGGSLLA
PGTLGSRSRSTTSFIRNVKPRTGEGLKNSRIPRNELLDILFKCFEDYEYWTLKGLREY
VKQPEVYLKEVLDSIAILNKRGPYALKYSLKPEYKGTMDAASVELRNQQASQSESSSI
DHTGKNTSPDNPGTNAEEDEDDDGVEMIDVV"

misc_feature complement(4185. .4202)
/gene="SPBC660.03c"
/note="ctaatgaattcatattag, splice branch and acceptor"

misc_feature complement(4282. .4287)
/gene="SPBC660.03c"
/note="gtatgt, splice donor sequence"

misc_feature complement(4432. .4447)
/gene="SPBC660.03c"
/note="ctaacttttttttcag, splice branch and acceptor"

misc_feature complement(4485. .4490)
/gene="SPBC660.03c"
/note="gtatgt, splice donor sequence"

gene complement(6039. .7082)

```
                /gene="SPBC660.04c"
     CDS        complement(6039. .7082)
                /gene="SPBC660.04c"
                /note="SPBC660.04c, len:347; SPBC660.04c, len:347"
                /codon_start=1
                /label=fbp1
                /product="fructose-1,6-bisphosphatase"
                /protein_id="CAA22524.1"
                /db_xref="GI:4049503"
                /translation="MKKDLDEIDTDIVTLSSFILQEQRRYNQKHKNEEGKPCIIQEAS
                GELSLLLNSLQFSFKFIANTIRKAELVNLIGLSGIVNSTGDEQKKLDKICNDIFITAM
                KSNGCCKLIVSEEEEDLIVVDSNGSYAVTCDPIDGSSNIDAGVSVGTIFGIYKLRPGS
                QGDISDVLRPGKEMVAAGYTMYGASAHLLLTTGHRVNGFTLDTDIGEFILTHRNMKMP
                LQHSIYSINEGYTAFWDEKIARFIAHLKESTPDKKPYSARYIGSMVADMHRTILYGGL
                FAYPCSKGNNGKLRLLYECFPMAFLVEQAGGIAVNDKGDRILDLVPKTLHGKSSIWLG
                SKHEVEEYINFIK"
     misc_feature   complement(6066. .6833)
                /gene="SPBC660.04c"

  ry Match              8.8%;  Score 18;  DB 8;  Length 43325;
    Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     175 AAGAAAACATCTACTTTG 192
           ||||||||||||||||||
Db   22202 AAGAAAACATCTACTTTG 22219


RESULT  11
AC022747/c
LOCUS       AC022747    83536 bp    DNA             HTG       06-FEB-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-131K9 map 4, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC022747
VERSION     AC022747.1  GI:6987626
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 83536)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-131K9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 83536)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu

-------------- Project Information
    Center project name: L5770
    Center clone name: 131_K_9
--------------.
* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      918: contig of 918 bp in length
*                    gap of unknown length
*      919     1824: contig of 906 bp in length
*                    gap of unknown length
*     1825     2752: contig of 928 bp in length
*                    gap of unknown length
*     2753     3669: contig of 917 bp in length
*                    gap of unknown length
*     3670     4548: contig of 879 bp in length
*                    gap of unknown length
*     4549     5462: contig of 914 bp in length
*                    gap of unknown length
*     5463     6358: contig of 896 bp in length
*                    gap of unknown length
*     6359     7273: contig of 915 bp in length
*                    gap of unknown length
*     7274     8172: contig of 899 bp in length
*                    gap of unknown length
*     8173     9067: contig of 895 bp in length
*                    gap of unknown length
*     9068    10000: contig of 933 bp in length
*                    gap of unknown length
*    10001    10900: contig of 900 bp in length
*                    gap of unknown length
*    10901    11815: contig of 915 bp in length
*                    gap of unknown length
*    11816    12726: contig of 911 bp in length
*                    gap of unknown length
*    12727    13646: contig of 920 bp in length
*                    gap of unknown length
*    13647    14558: contig of 912 bp in length
*                    gap of unknown length
*    14559    15451: contig of 893 bp in length
*                    gap of unknown length
*    15452    16393: contig of 942 bp in length
*                    gap of unknown length
*    16394    17255: contig of 862 bp in length
*                    gap of unknown length
*    17256    18153: contig of 898 bp in length
*                    gap of unknown length
*    18154    19033: contig of 880 bp in length
*                    gap of unknown length
*    19034    19966: contig of 933 bp in length
*                    gap of unknown length
*    19967    20895: contig of 929 bp in length
*                    gap of unknown length
*    20896    21813: contig of 918 bp in length
*                    gap of unknown length
*    21814    22713: contig of 900 bp in length
*                    gap of unknown length
*    22714    23608: contig of 895 bp in length
*                    gap of unknown length
*    23609    24519: contig of 911 bp in length
*                    gap of unknown length
*    24520    25434: contig of 915 bp in length
*                    gap of unknown length
*    25435    26309: contig of 875 bp in length
*                    gap of unknown length
```

```
*    26310    27226: contig of 917 bp in length
*                 gap of unknown length
*    27227    28127: contig of 901 bp in length
*                 gap of unknown length
*    28128    29027: contig of 900 bp in length
*                 gap of unknown length
*    29028    29943: contig of 916 bp in length
*                 gap of unknown length
*    29944    30848: contig of 905 bp in length
*                 gap of unknown length
*    30849    31764: contig of 916 bp in length
*                 gap of unknown length
*    31765    32673: contig of 909 bp in length
*                 gap of unknown length
*    32674    33584: contig of 911 bp in length
*                 gap of unknown length
*    33585    34503: contig of 919 bp in length
*                 gap of unknown length
*    34504    35390: contig of 887 bp in length
*                 gap of unknown length
*    35391    36287: contig of 897 bp in length
*                 gap of unknown length
*    36288    37171: contig of 884 bp in length
*                 gap of unknown length
*    37172    38102: contig of 931 bp in length
*                 gap of unknown length
*    38103    38995: contig of 893 bp in length
*                 gap of unknown length
*    38996    39916: contig of 921 bp in length
*                 gap of unknown length
*    39917    40854: contig of 938 bp in length
*                 gap of unknown length
*    40855    41771: contig of 917 bp in length
*                 gap of unknown length
*    41772    42674: contig of 903 bp in length
*                 gap of unknown length
*    42675    43522: contig of 848 bp in length
*                 gap of unknown length
*    43523    44459: contig of 937 bp in length
*                 gap of unknown length
*    44460    45411: contig of 952 bp in length
*                 gap of unknown length
*    45412    46349: contig of 938 bp in length
*                 gap of unknown length
*    46350    47276: contig of 927 bp in length
*                 gap of unknown length
*    47277    48173: contig of 897 bp in length
*                 gap of unknown length
*    48174    49064: contig of 891 bp in length
*                 gap of unknown length
*    49065    49947: contig of 883 bp in length
*                 gap of unknown length
*    49948    50850: contig of 903 bp in length
*                 gap of unknown length
*    50851    51768: contig of 918 bp in length
*                 gap of unknown length
*    51769    52695: contig of 927 bp in length
*                 gap of unknown length
*    52696    53602: contig of 907 bp in length
*                 gap of unknown length
*    53603    54527: contig of 925 bp in length
*                 gap of unknown length
*    54528    55444: contig of 917 bp in length
*                 gap of unknown length
*    55445    56353: contig of 909 bp in length
*                 gap of unknown length
*    56354    57264: contig of 911 bp in length
*                 gap of unknown length
*    57265    58154: contig of 890 bp in length
*                 gap of unknown length
*    58155    59076: contig of 922 bp in length
*                 gap of unknown length
*    59077    60005: contig of 929 bp in length
*                 gap of unknown length
*    60006    60897: contig of 892 bp in length
*                 gap of unknown length
*    60898    61798: contig of 901 bp in length
*                 gap of unknown length
*    61799    62706: contig of 908 bp in length
*                 gap of unknown length
*    62707    63615: contig of 909 bp in length
*                 gap of unknown length
*    63616    64547: contig of 932 bp in length
*                 gap of unknown length

  Query Match             8.8%;  Score 18;  DB 51;  Length 83536;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     164 TTCTCTCTAATAAGAAAA 181
           ||||||||||||||||||
Db   60164 TTCTCTCTAATAAGAAAA 60147


RESULT  12
AL136089
LOCUS       AL136089   102995 bp    DNA             HTG       08-APR-2000
DEFINITION  Homo sapiens chromosome 6 clone RP1-278E11, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL136089
VERSION     AL136089.9  GI:7530184
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102995)
  AUTHORS   Smalley,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2000 this sequence version replaced gi:7320935.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00643  Length: 7736bp
            Contig_ID: 00773  Length: 17572bp
            Contig_ID: 00923  Length: 76087bp.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     7736: contig of 7736 bp in length
            *    7737 8536: gap of      800 bp
            *    8537    26108: contig of 17572 bp in length
            *   26109 26908: gap of      800 bp
            *   26909   102995: contig of 76087 bp in length.
FEATURES             Location/Qualifiers
     source          1. .102995
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP1-278E11"
                     /clone_lib="RPCI-1"
BASE COUNT    28381 a  22682 c  22914 g  27418 t   1600 others
ORIGIN
```

```
  Query Match              8.8%;  Score 18;  DB 40;  Length 102995;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    188 CTTTGAAACATCTACTGG 205
          ||||||||||||||||||
Db  77128 CTTTGAAACATCTACTGG 77145


RESULT 13
HS86F14/c
LOCUS       HS86F14    106571 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from PAC 86F14 on chromosome 1q23-1q24. Contains
            coagulation factor V, ESTs and STS.
ACCESSION   Z99572
VERSION     Z99572.1  GI:2769646
KEYWORDS    1q23-1q24; blood coagulation factor; factor V.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106571)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome 1 Project Group
            (http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 13, 1998 this sequence version replaced gi:2578147.
            IMPORTANT: This sequence is not the entire insert of clone 86F14.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre chromosome 1
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true right end of clone 206D15 is at 104.
            The true right end of clone 86F14 is at 106571.
            86F14 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .106571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q23-1q24"
                     /clone="RP1-86F14"
                     /clone_lib="RPCI-1"
     repeat_region   810. .1090
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   1270. .1360
                     /note="MLT1E repeat: matches 1. .84 of consensus"
     repeat_region   1410. .1587
                     /note="MLT1D repeat: matches 139. .323 of consensus"
     repeat_region   1575. .1858
                     /note="MLT1E repeat: matches 265. .568 of consensus"
     repeat_region   1876. .2745
                     /note="L1MB5 repeat: matches 921. .1 of consensus"
     repeat_region   2597. .2793
                     /note="L1 repeat: matches 5390. .5194 of consensus"
     repeat_region   2854. .3352
                     /note="L1ME1 repeat: matches 687. .170 of consensus"
     repeat_region   2971. .3352
                     /note="L1ME2 repeat: matches 570. .170 of consensus"
     repeat_region   3327. .3559
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   3580. .3873
                     /note="AluSc repeat: matches 1. .299 of consensus"
     repeat_region   3877. .4060
                     /note="L1MC2 repeat: matches 193. .10 of consensus"
     repeat_region   6468. .6575
                     /note="MIR2 repeat: matches 22. .145 of consensus"
     repeat_region   7024. .7238
                     /note="AluY repeat: matches 84. .299 of consensus;
                     incomplete repeat"
     repeat_region   7834. .8126
                     /note="AluSc repeat: matches 2. .295 of consensus"
     repeat_region   10054. .10223
                     /note="MIR repeat: matches 89. .262 of consensus"
     misc_feature    complement(10749. .11106)
                     /note="match: STS G05144"
     repeat_region   11725. .12081
                     /note="THE1C repeat: matches 371. .1 of consensus"
     repeat_region   12196. .12497
                     /note="AluSq repeat: matches 303. .1 of consensus"
     repeat_region   13253. .13367
                     /note="MIR2 repeat: matches 4. .127 of consensus"
     repeat_region   13398. .15749
                     /note="L1 repeat: matches 2577. .4940 of consensus"
     repeat_region   16076. .16554
                     /note="L1 repeat: matches 4910. .5390 of consensus"
     repeat_region   16409. .17294
                     /note="L1PA2 repeat: matches 1. .891 of consensus"
     repeat_region   17808. .19825
                     /note="L1 repeat: matches 3338. .5390 of consensus"
     repeat_region   19683. .20571
                     /note="L1PA6 repeat: matches 1. .890 of consensus"
     repeat_region   22732. .22827
                     /note="L1MB4 repeat: matches 797. .902 of consensus"
     repeat_region   22748. .22851
                     /note="L1MB5 repeat: matches 812. .922 of consensus"
     prim_transcript <25068. .25543
                     /note="match: H61071 H69565"
     repeat_region   25302. .25381
                     /note="MIR repeat: matches 49. .131 of consensus"
     repeat_region   25527. .25703
                     /note="MER45 repeat: matches 1. .178 of consensus"
     repeat_region   25932. .26062
                     /note="MLT1F repeat: matches 18. .138 of consensus"
     repeat_region   26055. .26264
                     /note="L1 repeat: matches 2971. .2759 of consensus"
     repeat_region   26266. .26378
                     /note="L1MA7 repeat: matches 1017. .895 of consensus"
     prim_transcript 27046. .>27579
                     /note="match: multiple ESTs; match: R71060 H69028 R82016
                     R10102 R82280 R82281; match: R82066 H69792 H79486 AA506861
                     T78091; match: W03874 H74282 D85329"
     CDS             complement(join(27431. .27577,28562. .28744,31530. .31681,
                     33638. .33782,36315. .36470,36919. .37022,37955. .38026,
                     39019. .39135,41033. .41212,42726. .42936,43904. .44140,
                     49624. .49798,53412. .56232,57414. .57626,59560. .59710,
                     62919. .63133,63758. .63857,65675. .65852,68300. .68465,
                     69764. .69985,72271. .72414,73672. .73884,85339. .85461,
                     95549. .95640,99347. .99504))
                     /codon_start=1
                     /product="factor V"
                     /protein_id="CAB16748.1"
                     /db_xref="GI:2769647"
                     /db_xref="SPTREMBL:O43737"
                     /translation="MFPGCPRLWVLVVLGTSWVGWGSQGTEAAQLRQFYVAAQGISWS
                     YRPEPTNSSLNLSVTSFKKIVYREYEPYFKKEKPQSTISGLLGPTLYAEVGDIIKVHF
                     KNKADKPLSIHPQGIRYSKLSEGASYLDHTFPAEKMDDAVAPGREYTYEWSISEDSGP
```

```
                     THDDPPCLTHIYYSHENLIEDFNSGLIGPLLICKKGTLTEGGTQKTFDKQIVLLFAVF
                     DESKSWSQSSSLMYTVNGYVNGTMPDITVCAHDHISWHLLGMSSGPELFSIHFNGQVL
                     EQNHHKVSAITLVSATSTTANMTVGPEGKWIISSLTPKHLQAGMQAYIDIKNCPKKTR
                     NLKKITREQRRHMKRWEYFIAAEEVIWDYAPVIPANMDKKYRSQHLDNFSNQIGKHYK
                     KVMYTQYEDESFTKHTVNPNMKEDGILGPIIRAQVRDTLKIVFKNMASRPYSIYPHGV
                     TFSPYEDEVNSSFTSGRNNTMIRAVQPGETYTYKWNILEFDEPTENDAQCLTRPYYSD
                     VDIMRDIASGLIGLLLICKSRSLDRQGIQRAADIEQQAVFAVFDENKSWYLEDNINKF
                     CENPDEVKRDDPKFYESNIMSTINGYVPESITTLGFCFDDTVQWHFCSVGTQNEILTI
                     HFTGHSFIYGKRHEDTLTLFPMRGESVTVTMDNVGTWMLTSMNSSPRSKKLRLKFRDV
                     KCIPDDDEDSYEIFEPPESTVMATRKMHDRLEPEDEESDADYDYQNRLAAALGIRSFR
                     NSSLNQEEEEFNLTALALENGTEFVSSNTDIIVGSNYSSPSNISKFTVNNLAEPQKAP
                     SHQQATTAGSPLRHLIGKNSVLNSSTAEHSSPYSEDPIEDPLQPDVTGIRLLSLGAGE
                     FKSQEHAKHKGPKVERDQAAKHRFSWMKLLAHKVGRHLSQDTGSPSGMRPWEDLPSQD
                     TGSPSRMRPWKDPPSDLLLLKQSNSSKILVGRWHLASEKGSYEIIQDTDEDTAVNNWL
                     ISPQNASRAWGESTPLANKPGKQSGHPKFPRVRHKSLQVRQDGGKSRLKKSQFLIKTR
                     KKKKEKHTHHAPLSPRTFHPLRSEAYNTFSERRLKHSLVLHKSNETSLPTDLNQTLPS
                     MDFGWIASLPDHNQNSSNDTGQASCPPGLYQTVPPEEHYQTFPIQDPDQMHSTSDPSH
                     RSSSPELSEMLEYDRSHKSFPTDISQMSPSSEHEVWQTVISPDLSQVTLSPELSQTNL
                     SPDLSHTTLSPELIQRNLSPALGQMPISPDLSHTTLSPDLSHTTLSLDLSQTNLSPEL
                     SQTNLSPALGQMPLSPDLSHTTLSLDFSQTNLSPELSHMTLSPELSQTNLSPALGQMP
                     ISPDLSHTTLSLDFSQTNLSPELSQTNLSPALGQMPLSPDPSHTTLSLDLSQTNLSPE
                     LSQTNLSPDLSEMPLFADLSQIPLTPDLDQMTLSPDLGETDLSPNFGQMSLSPDLSQV
                     TLSPDISDTTLLPDLSQISPPPDLDQIFYPSESSQSLLLQEFNESFPYPDLGQMPSPS
                     SPTLNDTFLSKEFNPLVIVGLSKDGTDYIEIIPKEEVQSSEDDYAEIDYVPYDDPYKT
                     DVRTNINSSRDPDNIAAWYLRSNNGNRRNYYIAAEEISWDYSEFVQRETDIEDSDDIP
                     EDTTYKKVVFRKYLDSTFTKRDPRGEYEEHLGILGPIIRAEVDDVIQVRFKNLASRPY
                     SLHAHGLSYEKSSEGKTYEDDSPEWFKEDNAVQPNSSYTYVWHATERSGPESPGSACR
                     AWAYYSAVNPEKDIHSGLIGPLLICQKGILHKDSNMPMDMREFVLLFMTFDEKKSWYY
                     EKKSRSSWRLTSSEMKKSHEFHAINGMIYSLPGLKMYEQEWVRLHLLNIGGSQDIHVV
                     HFHGQTLLENGNKQHQLGVWPLLPGSFKTLEMKASKPGWWLLNTEVGENQRAGMQTPF
                     LIMDRDCRMPMGLSTGIISDSQIKASEFLGYWEPRLARLNNGGSYNAWSVEKLAAEFA
                     SKPWIQVDMQKEVIITGIQTQGAKHYLKSCYTTEFYVAYSSNQINWQIFKGNSTRNVM
                     YFNGNSDASTIKENQFDPPIVARYIRISPTRAYNRPTLRLELQGCEVNGCSTPLGMEN
                     GKIENKQITASSFKKSWWGDYWEPFRARLNAQGRVNAWQAKANNNKQWLEIDLLKIKK
                     ITAIITQGCKSLSSEMYVKSYTIHYSEQGVEWKPYRLKSSMVDKIFEGNTNTKGHVKN
                     FFNPPIISRFIRVIPKTWNQSIALRLELFGCDIY"
     repeat_region   27629. .27788
                     /note="AluJo repeat: matches 302. .140 of consensus;
                     incomplete repeat"
     repeat_region   28002. .28176
                     /note="MIR repeat: matches 75. .259 of consensus"
     repeat_region   29041. .29084
                     /note="MIR2 repeat: matches 103. .146 of consensus"
     repeat_region   31903. .31988
                     /note="MIR repeat: matches 104. .17 of consensus"
     repeat_region   34738. .35036
                     /note="AluSc repeat: matches 299. .1 of consensus"
     repeat_region   35051. .35542
                     /note="L1 repeat: matches 3215. .3729 of consensus"
     repeat_region   36628. .36808
                     /note="MIR repeat: matches 40. .235 of consensus"
     repeat_region   38333. .38451
                     /note="AluJo repeat: matches 1. .134 of consensus;
                     incomplete repeat"
     repeat_region   40215. .40272
                     /note="MADE1 repeat: matches 1. .53 of consensus"
     repeat_region   43274. .43327
                     /note="MIR2 repeat: matches 146. .93 of consensus"
     repeat_region   43456. .43548
                     /note="L1MA8 repeat: matches 1038. .944 of consensus"
     repeat_region   44951. .46774
                     /note="L1 repeat: matches 3613. .5390 of consensus"
     repeat_region   46627. .46980
                     /note="L1MA7 repeat: matches 2. .364 of consensus"
     repeat_region   46983. .47446

  Query Match             8.8%;  Score 18;  DB 10;  Length 106571;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     171 TAATAAGAAAACATCTAC 188
           ||||||||||||||||||
Db    1250 TAATAAGAAAACATCTAC 1233
```

```
RESULT  14
AC010429
LOCUS       AC010429   133783 bp    DNA             HTG       05-APR-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2199L14, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC010429
VERSION     AC010429.3  GI:7417548
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133783)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133783)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:7212886.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------Summary Statistics
            Consensus quality: 132291 bases at least Q40
            Consensus quality: 133603 bases at least Q30
            Consensus quality: 133733 bases at least Q20
            Estimated insert size: 133783; sum-of-contigs estimation
            Estimated insert size: 233000; pulse field gel estimation
            Quality coverage:  4.87x in Q20 bases; pulse field gel estimation
            Quality coverage:  8.48x in Q20 bases; sum-of-contigs estimation
            -------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9662: contig of 9662 bp in length
            *                   gap of unknown length
            *     9663   133783: contig of 124121 bp in length.
FEATURES              Location/Qualifiers
     source           1. .133783
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="5"
                      /clone="CTD-2199L14"
BASE COUNT    43496 a  25289 c  24707 g  40291 t
ORIGIN


  Query Match             8.8%;  Score 18;  DB 72;  Length 133783;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     178 AAAACATCTACTTTGAAA 195
           ||||||||||||||||||
Db   62087 AAAACATCTACTTTGAAA 62104


RESULT  15
AP000817/c
LOCUS       AP000817   139740 bp    DNA             HTG       18-FEB-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-21K9 map 11q22, WORKING DRAFT
            SEQUENCE, 25 unordered pieces.
ACCESSION   AP000817
```

```
VERSION     AP000817.2  GI:7007459
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:CMB9-21K9.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139740)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 139,740 genomic DNA of 11q22
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 139740)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Feb 19, 2000 this sequence version replaced gi:6997652.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.RIKEN.go.jp/
                Contact: hattori@gsc.RIKEN.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-21K9
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 112677 bases at least Q40
                Consensus quality: 121194 bases at least Q30
                Consensus quality: 125720 bases at least Q20
                Insert size: 129024; sum-of-contigs
                Quality coverage: 4.32x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            25 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     15833 contig of  15833 bp in length
              16334     34080 contig of  17747 bp in length
              34581     46531 contig of  11951 bp in length
              47032     56604 contig of   9573 bp in length
              57105     65754 contig of   8650 bp in length
              66255     71355 contig of   5101 bp in length
              71856     77672 contig of   5817 bp in length
              78173     81813 contig of   3641 bp in length
              82314     87385 contig of   5072 bp in length
              87886     92261 contig of   4376 bp in length
              92762     97060 contig of   4299 bp in length
              97561    102247 contig of   4687 bp in length
             102748    106863 contig of   4116 bp in length
             107364    109944 contig of   2581 bp in length
             110445    113790 contig of   3346 bp in length
             114291    117299 contig of   3009 bp in length
             117800    121336 contig of   3537 bp in length
             121837    125424 contig of   3588 bp in length
             125925    128032 contig of   2108 bp in length
             128533    130201 contig of   1669 bp in length
             130702    131963 contig of   1262 bp in length
             132464    134276 contig of   1813 bp in length
             134777    135777 contig of   1001 bp in length
             136278    137928 contig of   1651 bp in length
             138429    139740 contig of   1312 bp in length
            Sequence updated (16-Feb-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    15833: contig of 15833 bp in length
            *   15834 16333: gap of      500 bp
            *   16334    34076: contig of 17743 bp in length
            *   34077 34581: gap of      505 bp
            *   34582    46531: contig of 11950 bp in length
            *   46532 47034: gap of      503 bp
            *   47035    56602: contig of 9568 bp in length
            *   56603 57107: gap of      505 bp
            *   57108    65754: contig of 8647 bp in length
            *   65755 66255: gap of      501 bp
            *   66256    71355: contig of 5100 bp in length
            *   71356 71866: gap of      511 bp
            *   71867    77672: contig of 5806 bp in length
            *   77673 78174: gap of      502 bp
            *   78175    81813: contig of 3639 bp in length
            *   81814 82313: gap of      500 bp
            *   82314    87384: contig of 5071 bp in length
            *   87385 87886: gap of      502 bp
            *   87887    92260: contig of 4374 bp in length
            *   92261 92762: gap of      502 bp
            *   92763    97056: contig of 4294 bp in length
            *   97057 97566: gap of      510 bp
            *   97567   102247: contig of 4681 bp in length
            *  102248 102760: gap of      513 bp
            *  102761   106861: contig of 4101 bp in length
            *  106862 107367: gap of      506 bp
            *  107368   109928: contig of 2561 bp in length
            *  109929 110444: gap of      516 bp
            *  110445   113790: contig of 3346 bp in length
            *  113791 114290: gap of      500 bp
            *  114291   117299: contig of 3009 bp in length
            *  117300 117801: gap of      502 bp
            *  117802   121336: contig of 3535 bp in length
            *  121337 121836: gap of      500 bp
            *  121837   125424: contig of 3588 bp in length
            *  125425 125924: gap of      500 bp
            *  125925   128031: contig of 2107 bp in length
            *  128032 128536: gap of      505 bp
            *  128537   130196: contig of 1660 bp in length
            *  130197 130709: gap of      513 bp
            *  130710   131962: contig of 1253 bp in length
            *  131963 132465: gap of      503 bp
            *  132466   134274: contig of 1809 bp in length
            *  134275 134778: gap of      504 bp
            *  134779   135777: contig of 999 bp in length
            *  135778 136277: gap of      500 bp
            *  136278   137927: contig of 1650 bp in length
            *  137928 138429: gap of      502 bp
            *  138430   139740: contig of 1311 bp in length.
FEATURES             Location/Qualifiers
     source          1. .139740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="CMB9-21K9"
                     /map="11q22"
BASE COUNT    40055 a  21062 c  21342 g  43261 t  14020 others
ORIGIN


  Query Match              8.8%;  Score 18;  DB 31;  Length 139740;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 TGCTTTTCTCTCTAATAA 176
```

```
                  ||||||||||||||||||
Db  94300 TGCTTTTCTCTCTAATAA 94283
```

Search completed: October 3, 2000, 12:59:51
Job time: 9393 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: October 3, 2000, 10:23:18 ; Search time 1545.92 Seconds
(without alignments)
247.054 Million cell updates/sec

Title: US-09-065-672-1
Perfect score: 214
Sequence: 1 CTAAGGCGTGCAAACAGAGC.....TACTTTGAAACATCTACTGG 214

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

size : 0

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*
44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_pr5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_vi1:*
82: gb_vi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19	8.9	143068	11	HSU95626	U95626 Homo sapien
	2	19	8.9	216514	55	AC018744	AC018744 Oryza sat
c	3	18	8.4	2698	7	YSCNACT	M23166 S.cerevisia
c	4	18	8.4	2699	5	I08122	I08122 Sequence 1
c	5	18	8.4	2724	5	I09397	I09397 Sequence 5
c	6	18	8.4	3347	7	SCNAT	X15135 Yeast NAT 1
	7	18	8.4	3530	7	SCYDL040C	Z74088 S.cerevisia
	8	18	8.4	36687	7	SCCIVL37K	Z71781 S.cerevisia
	9	18	8.4	43325	8	SPBC660	AL034563 S.pombe c
c	10	18	8.4	83536	51	AC022747	AC022747 Homo sapi
	11	18	8.4	102995	40	AL136089	AL136089 Homo sapi
c	12	18	8.4	106571	10	HS86F14	Z99572 Human DNA s
	13	18	8.4	133783	72	AC010429	AC010429 Homo sapi
c	14	18	8.4	139740	31	AP000817	AP000817 Homo sapi
	15	18	8.4	141107	67	AC022414	AC022414 Homo sapi
c	16	18	8.4	145342	69	AC023220	AC023220 Homo sapi
c	17	18	8.4	151071	31	AP001795	AP001795 Homo sapi
c	18	18	8.4	154208	78	AC021203	AC021203 Homo sapi
c	19	18	8.4	158097	54	AC008471	AC008471 Homo sapi
	20	18	8.4	159624	56	AC011021	AC011021 Homo sapi
	21	18	8.4	161624	54	AC011640	AC011640 Homo sapi
	22	18	8.4	171300	43	AC021986	AC021986 Homo sapi
	23	18	8.4	178071	67	AC024177	AC024177 Homo sapi
c	24	18	8.4	182341	32	AL139238	AL139238 Homo sapi

	25	18	8.4	182482	43	AC016703	AC016703 Homo sapi
	26	18	8.4	186243	52	AC012022	AC012022 Homo sapi
c	27	18	8.4	195832	78	AC019184	AC019184 Homo sapi
c	28	18	8.4	216215	10	HSG256O22	AL080239 Human DNA
c	29	18	8.4	240327	69	AC022422	AC022422 Homo sapi
	30	18	8.4	260270	40	AL135840	AL135840 Homo sapi
	31	17	7.9	526	13	G61963	G61963 SHGC-89514
	32	17	7.9	1155	39	AF100634	AF100634 Homo sapi
	33	17	7.9	1795	9	HSY14873	Y14873 Homo sapien
c	34	17	7.9	1850	1	ECEXPIR	X96440 E.chrysanth
	35	17	7.9	2415	9	AK001422	AK001422 Homo sapi
	36	17	7.9	2489	9	HSDARC	X85785 H.sapiens D
	37	17	7.9	2772	11	AF055992	AF055992 Homo sapi
c	38	17	7.9	2795	11	HSU43899	U43899 Human signa
	39	17	7.9	3068	10	S76830	S76830 glycoprotei
c	40	17	7.9	13271	2	AE001168	AE001168 Borrelia
	41	17	7.9	23332	42	AC014464	AC014464 Drosophil
c	42	17	7.9	23379	34	CELT08E11	AF077546 Caenorhab
	43	17	7.9	36589	9	AP001049	AP001049 Homo sapi
[redacted]		17	7.9	39752	9	D86993	D86993 Homo sapien
[redacted]		17	7.9	40677	50	HSAC000099	AC000099 Cosmid g0

ALIGNMENTS

RESULT 1
HSU95626/c
LOCUS HSU95626 143068 bp DNA PRI 16-MAY-1997
DEFINITION Homo sapiens ccr2b (ccr2), ccr2a (ccr2), ccr5 (ccr5) and ccr6 (ccr6) genes, complete cds, and lactoferrin (lactoferrin) gene, partial cds, complete sequence.
ACCESSION U95626
VERSION U95626.1 GI:2104517
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143068)
AUTHORS McCombie,W.R., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D., Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K., Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J., Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M., Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E., Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y., Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
TITLE Human BAC clone 110P12
[redacted]RNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 143068)
AUTHORS McCombie,R.W., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D., Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K., Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J., Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M., Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E., Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y., Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1997) Advanced Genome Sequence Analysis Course, Cold Spring Harbor Laboratory, 1Bungtown Rd., Cold Spring Harbor, NY 11724, USA
COMMENT Regions with single-strand coverage are as follows:

31434 - 31443 37900 - 37968 53303 - 53357
59166 - 59206 63708 - 63998 65200 - 65335
78605 - 78713 92135 - 92137 112377 - 112551
112643 - 112778 134284 - 134309 134914 - 135019
143046 - 144068.

FEATURES Location/Qualifiers
source 1. .143068
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="BAC 110P12"
mRNA 46056. .47997
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced mRNA encoding B-form carboxyl tail. Accession Number: U80924."
/product="ccr2b"
mRNA join(46056. .47046,48255. .49505)
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced mRNA encoding A-form carboxyl tail, Accession Number U80924."
/product="ccr2a"
gene 46056. .49505
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) mRNA (Accession Number U80924), two alternatively spliced mRNAs."
CDS join(46106. .47046,48255. .48438)
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced A-form, Encoded by GenBank Accession Number U80924, gi 1168965"
/codon_start=1
/product="ccr2a"
/protein_id="AAB57791.1"
/db_xref="GI:2104518"
/translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRSLFHIALGCRIAPLQKPVCGG
PGVRPGKNVKVTTQGLLDGRGKGKSIGRAPEASLQDKEGA"
CDS 46106. .47188
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2), Accession Number: 1168965"
/codon_start=1
/product="ccr2b"
/protein_id="AAB57792.1"
/db_xref="GI:2104519"
/translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRRYLSVFFRKHITKRFCKQCPV
FYRETVDGVTSTNTPSTGEQEVSAGL"
mRNA join(59531. .59573,61472. .64785)
/gene="ccr5"
/note="confirmed by similarity to Human cc chemokine receptor 5 (ccr5) mRNA. Accession number: U54994."
/product="ccr5"
gene 59531. .64785
/gene="ccr5"
CDS 61483. .62541
/gene="ccr5"
/note="confirmed by similarity to human CC chemokine receptor 5 (ccr5) protein, encoded by GenBank Accession Number U54994, gi 1457946"
/codon_start=1
/product="ccr5"
/protein_id="AAB57793.1"
/db_xref="GI:2104520"
/translation="MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFG
FVGNMLVILILINCKRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQ
LLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFAS
LPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVICYSGIL

```
                     KTLLRCRNEKKRHRAVRLIFTIMIVYFLFWAPYNIVLLLNTFQEFFGLNNCSSSNRLD
                     QAMQVTETLGMTHCCINPIIYAFVGEKFRNYLLVFFQKHIAKRFCKCCSIFQQEAPER
                     ASSVYTRSTGEQEISVGL"
     mRNA            96634. .97683
                     /note="Putative mRNA identified by homology to CCR5 mRNA."
                     /product="ccr6"
                     /evidence=not_experimental
     gene            96642. .97676
                     /gene="ccr6"
     CDS             96642. .97676
                     /gene="ccr6"
                     /note="Identified as a gene by Grail Version 1.3c.
                     Translated sequence exhibits 42% sequence identity to CCR5
                     protein."
                     /codon_start=1
                     /evidence=not_experimental
                     /product="ccr6"
                     /protein_id="AAB57794.1"
                     /db_xref="GI:2104521"
                     /translation="MANYTLAPEDEYDVLIEGELESDEAEQCDKYDAQALSAQLVPSL
                     CSAVFVIGVLDNLLVVLILVKYKGLKRVENIYLLNLAVSNLCFLLTLPFWAHAGGDPM
                     CKILIGLYFVGLYSETFFNCLLTVQRYLVFLHKGNFFSARRRVPCGIITSVLAWVTAI
                     LATLPEYVVYKPQMEDQKYKCAFSRTPFLPADETFWKHFLTLKMNISVLVLPLFIFTF
                     LYVQMRKTLRFREQRYSLFKLVFAIMVVFLLMWAPYNIAFFLSTFKEHFSLSDCKSSY
                     NLDKSVHITKLIATTHCCINPLLYAFLDGTFSKYLCRCFHLRSNTPLQPRGQSAQGTS
                     REEPDHSTEV"
     mRNA            complement(join(124605. .124816,126528. .126717,
                     127884. .128068,130006. .130073,132023. .132164,
                     133863. .134018,135022. .135075,135890. .135980,
                     137445. .137599,138436. .138610,139077. .>139255))
                     /gene="lactoferrin"
                     /note="confirmed by similarity to lactoferrin mRNA,
                     accession number M73700"
                     /product="lactoferrin"
     gene            complement(124605. .>139255)
                     /gene="lactoferrin"
     CDS             complement(join(124782. .124816,126528. .126717,
                     127884. .128068,130006. .130073,132023. .132164,
                     133863. .134018,135022. .135075,135890. .135980,
                     137445. .137599,138436. .138610,139077. .>139253))
                     /gene="lactoferrin"
                     /note="confirmed by similarity to lactoferrin protein,
                     encoded by GenBank  Accession Number M73700, gi  186818"
                     /codon_start=1
                     /product="lactoferrin"
                     /protein_id="AAB57795.1"
                     /db_xref="GI:2104522"
                     /translation="DLSDEAERDEYELLCPDNTRKPVDKFKDCHLARVPSHAVVARSV
                     NGKEDAIWNLLRQAQEKFGKDKSPKFQLFGSPSGQKDLLFKDSAIGFSRVPPRIDSGL
                     YLGSGYFTAIQNLRKSEEEVAARRARVVWCAVGEQELRKCNQWSGLSEGSVTCSSAST
                     TEDCIALVLKGEADAMSLDGGYVYTAGKCGLVPVLAENYKSQQSSDPDPNCVDRPVEG
                     YLAVAVVRRSDTSLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQTGSCKFDEYFSQS
                     CAPGSDPRSNLCALCIGDEQGENKCVPNSNERYYGYTGAFRCLAENAGDVAFVKDVTV
                     LQNTDGNNNDAWAKDLKLADFALLCLDGKRKPVTEARSCHLAMAPNHAVVSRMDKVER
                     LKQVLLHQQAKFGRNGSDCPDKFCLFQSETKNLLFNDNTECLARLHGKTTYEKYLGPQ
                     YVAGITNLKKCSTSPLLEACEFLRK"
BASE COUNT    41194 a  30122 c  32403 g  39349 t
ORIGIN


  Query Match             8.9%;  Score 19;  DB 11;  Length 143068;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     82 CTGGTCCCCACCTTTGCAG 100
          ||||||||||||||||||||
Db  32691 CTGGTCCCCACCTTTGCAG 32673


RESULT   2
AC018744
LOCUS       AC018744   216514 bp    DNA             HTG       07-MAR-2000
DEFINITION  Oryza sativa chromosome 10 clone 15022, *** SEQUENCING IN PROGRESS
            ***, 16 unordered pieces.
ACCESSION   AC018744
VERSION     AC018744.2  GI:7191023
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 216514)
  AUTHORS   McCombie,W.R.
  TITLE     Rice genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216514)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Mar 7, 2000 this sequence version replaced gi:6730690.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   158180: contig of 158180 bp in length
            *                   gap of unknown length
            *  158181   174277: contig of 16097 bp in length
            *                   gap of unknown length
            *  174278   188853: contig of 14576 bp in length
            *                   gap of unknown length
            *  188854   192653: contig of 3800 bp in length
            *                   gap of unknown length
            *  192654   196182: contig of 3529 bp in length
            *                   gap of unknown length
            *  196183   198852: contig of 2670 bp in length
            *                   gap of unknown length
            *  198853   201033: contig of 2181 bp in length
            *                   gap of unknown length
            *  201034   203123: contig of 2090 bp in length
            *                   gap of unknown length
            *  203124   205004: contig of 1881 bp in length
            *                   gap of unknown length
            *  205005   206840: contig of 1836 bp in length
            *                   gap of unknown length
            *  206841   208587: contig of 1747 bp in length
            *                   gap of unknown length
            *  208588   210229: contig of 1642 bp in length
            *                   gap of unknown length
            *  210230   211854: contig of 1625 bp in length
            *                   gap of unknown length
            *  211855   213466: contig of 1612 bp in length
            *                   gap of unknown length
            *  213467   215011: contig of 1545 bp in length
            *                   gap of unknown length
            *  215012   216514: contig of 1503 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216514
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /clone="15022"
BASE COUNT    62170 a  45887 c  47331 g  60900 t    226 others
ORIGIN


  Query Match             8.9%;  Score 19;  DB 55;  Length 216514;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    175 CTCTCTAATAAGAAAACAT 193
          ||||||||||||||||||
Db 137518 CTCTCTAATAAGAAAACAT 137536


RESULT   3
YSCNACT/c
LOCUS       YSCNACT      2698 bp    mRNA            PLN       16-FEB-1996
DEFINITION  S.cerevisiae N-acetyltransferase (AAA1) mRNA, complete cds.
ACCESSION   M23166 J04837
VERSION     M23166.1  GI:172027
KEYWORDS    N-acetyltransferase.
SOURCE      Saccharomyces cerevisiae (strain TD71.8) (clone: pBN9) cDNA to
            mRNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2698)
  AUTHORS   Lee,F.J., Lin,L.W. and Smith,J.A.
  TITLE     Molecular cloning and sequencing of a cDNA encoding N
            alpha-acetyltransferase from Saccharomyces cerevisiae
  JOURNAL   J. Biol. Chem. 264 (21), 12339-12343 (1989)
  MEDLINE   89308659
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            F.-J.Lee, 10-APR-1989.
FEATURES             Location/Qualifiers
     source          1. .2698
                     /organism="Saccharomyces cerevisiae"
                     /strain="TD71.8"
                     /db_xref="taxon:4932"
                     /clone="pBN9"
                     /note="(vector lambda gt11)"
     gene            1. .2698
                     /gene="NAT1"
                     /db_xref="SGD:S0002198"
     mRNA            <22. .2698
                     /gene="NAT1"
                     /note="N-acetyltransferase mRNA"
                     /db_xref="SGD:S0002198"
     CDS             22. .2586
                     /gene="NAT1"
                     /note="N-acetyltransferase"
                     /codon_start=1
                     /db_xref="SGD:S0002198"
                     /protein_id="AAA88728.1"
                     /db_xref="GI:172028"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD.
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
BASE COUNT      926 a    491 c    533 g    748 t
ORIGIN      Chromosome 4.


  Query Match             8.4%;  Score 18;  DB 7;  Length 2698;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db    765 TGCTTTTCTCTCTAATAA 748


RESULT   4
I08122/c
LOCUS       I08122       2699 bp                    PAT       02-DEC-1994
DEFINITION  Sequence 1 from Patent EP 0334004.
ACCESSION   I08122
VERSION     I08122.1  GI:589163
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2699)
  AUTHORS   Smith,J.A. and Lee,F-J.S.
  TITLE     Isolation, purification, characterization, cloning and sequencing
            of N alpha-acetyltransferase
  JOURNAL   Patent: EP 0334004-A1 1 27-SEP-1989;
FEATURES             Location/Qualifiers
     source          1. .2699
                     /organism="unknown"
BASE COUNT      927 a    492 c    532 g    748 t
ORIGIN


  Query Match             8.4%;  Score 18;  DB 5;  Length 2699;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db    765 TGCTTTTCTCTCTAATAA 748


RESULT   5
I09397/c
LOCUS       I09397       2724 bp                    PAT       02-DEC-1994
DEFINITION  Sequence 5 from Patent WO 8907138.
ACCESSION   I09397
VERSION     I09397.1  GI:587894
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2724)
  AUTHORS   Smith,J.A. and Lee,F-J.S.
  JOURNAL   Patent: WO 8907138-A 5 10-AUG-1989;
FEATURES             Location/Qualifiers
     source          1. .2724
                     /organism="unknown"
BASE COUNT      952 a    491 c    533 g    748 t
ORIGIN


  Query Match             8.4%;  Score 18;  DB 5;  Length 2724;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db    765 TGCTTTTCTCTCTAATAA 748


RESULT   6
SCNAT/c
LOCUS       SCNAT        3347 bp    DNA             PLN       12-SEP-1993
DEFINITION  Yeast NAT 1 gene for N-terminal acetyltransferase.
ACCESSION   X15135
VERSION     X15135.1  GI:4027
KEYWORDS    acetyltransferase; NAT 1 gene.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3347)
```

```
  AUTHORS   Grunstein,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1989) Grunstein M., UCLA, Biology Department, Los
            Angeles CA 90024, USA
REFERENCE   2  (bases 1 to 3347)
  AUTHORS   Mullen,J.R., Kayne,P.S., Moerschell,R.P., Tsunasawa,S.,
            Gribskov,M., Colavito-Shepanski,M., Grunstein,M., Sherman,F. and
            Sternglanz,R.
  TITLE     Identification and characterization of genes and mutants for an
            N-terminal acetyltransferase from yeast
  JOURNAL   EMBO J. 8 (7), 2067-2075 (1989)
  MEDLINE   90005412
COMMENT     See <X01419> for overlapping sequence.
FEATURES             Location/Qualifiers
     source          1. .3347
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
     CDS             337. .2901
                     /note="acetyltransferase (AA 1-854)"
                     /codon_start=1
                     /protein_id="CAA33233.1"
                     /db_xref="GI:4028"
                     /db_xref="SWISS-PROT:P12945"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
     CDS             3330. .>3347
                     /note="SIR2 (AA 1-6)"
                     /codon_start=1
                     /protein_id="CAA33234.1"
                     /db_xref="GI:578200"
                     /translation="MTIPHM"
BASE COUNT     1127 a    625 c    653 g    942 t
ORIGIN
```

```
Query Match             8.4%;  Score 18;  DB 7;  Length 3347;
Best Local Similarity   100.0%;  Pred. No. 29;
Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     168 TGCTTTTCTCTCTAATAA 185
           ||||||||||||||||||
Db    1080 TGCTTTTCTCTCTAATAA 1063
```

```
RESULT   7
SCYDL040C
LOCUS       SCYDL040C    3530 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome IV reading frame ORF YDL040c.
ACCESSION   Z74088 Z71256
VERSION     Z74088.1  GI:1431024
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3530)
  AUTHORS   Paulin,L., Saren,A.M. and Laamanen,P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3530)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1996) Data collected by MIPS on behalf of the
            European yeast chromosome IV sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .3530
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
     gene            complement(419. .2983)
                     /gene="NAT1"
                     /db_xref="SGD:S0002198"
     CDS             complement(419. .2983)
                     /gene="NAT1"
                     /note="ORF YDL040c"
                     /codon_start=1
                     /db_xref="SGD:S0002198"
                     /protein_id="CAA98599.1"
                     /db_xref="GI:1431025"
                     /db_xref="SWISS-PROT:P12945"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
BASE COUNT     1009 a    672 c    660 g    1189 t
ORIGIN
```

```
Query Match             8.4%;  Score 18;  DB 7;  Length 3530;
Best Local Similarity   100.0%;  Pred. No. 29;
Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     168 TGCTTTTCTCTCTAATAA 185
           ||||||||||||||||||
Db    2240 TGCTTTTCTCTCTAATAA 2257
```

```
RESULT   8
SCCIVL37K
LOCUS       SCCIVL37K   36687 bp    DNA             PLN       14-FEB-1997
DEFINITION  S.cerevisiae chromosome IV left arm (EU) DNA segment (36687 bp).
ACCESSION   Z71781
VERSION     Z71781.1  GI:1279667
KEYWORDS    act2 gene; actin; FAD synthetase; FAD1 gene; MPS1 gene; NAM1 gene;
            NAT 1 gene; protein kinase; protein phosphatase; SIR2 gene; SIT4
            gene.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 36687)
  AUTHORS   Saren,A.M., Laamanen,P., Lejarcegui,J.B. and Paulin,L.
  TITLE     The sequence of a 36.7 kb segment on the left arm of chromosome IV
            from Saccharomyces cerevisiae reveals 20 non-overlapping open
            reading frames (ORFs) including SIT4, FAD1, NAM1, RNA11, SIR2,
            NAT1, PRP9, ACT2 and MPS1 and 11 new ORFs
  JOURNAL   Yeast 13 (1), 65-71 (1997)
  MEDLINE   97197972
REFERENCE   2  (bases 1 to 36687)
  AUTHORS   Paulin,L.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (23-APR-1996) Paulin L., Institute of Biotechnology, DNA
            sequencing & Synthesis Laboratory, Biocentre 1, P.O.BOX 56
            (Viikinkaari 9), FIN-00014 University of Helsinki, Finland
FEATURES             Location/Qualifiers
     source          1. .36687
                     /organism="Saccharomyces cerevisiae"
                     /strain="alpha S288c"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
                     /clone="cosmid 2G24"
                     /clone_lib="chromosome library of C. Jacq (Paris)"
     CDS             complement(<1. .556)
                     /standard_name="D2690"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96441.1"
                     /db_xref="GI:1279668"
                     /db_xref="SPTREMBL:Q07351"
                     /translation="MLVSSSFASSIDSVMMLVSSSFASSIDSVMSHETMSLRRNPPFI
                     DTPEKMPNPTASPNGTIHHLIDPSLPLLSSTTSSSRSTLSSTLNSPPPPPLTTSYSSY
                     NSSACQSITSSPTDNTALAHNSKCYFPHSLSPTPLSSNSSSHVILPPISSFTNLITVA
                     EREFNGRSNSLHANFTSPVPRTVLD"
     gene            2114. .3049
                     /gene="SIT4"
                     /db_xref="SGD:S0002205"
     CDS             2114. .3049
                     /gene="SIT4"
                     /standard_name="D2693"
                     /note="Acc.no. M24395"
                     /codon_start=1
                     /db_xref="SGD:S0002205"
                     /product="protein phosphatase catalytic subunit homologue
                     SIT4"
                     /protein_id="CAA96442.1"
                     /db_xref="GI:1279669"
                     /db_xref="SWISS-PROT:P20604"
                     /translation="MVSRGPDEWLETIKKCQALTENEMKQLCEMVKELLMEESNIQPV
                     QTPVTVCGDIHGQFHDLLELFRTAGGFPDDINYIFLGDYVDRGYYSLETFTLLMCLKV
                     KYPAKITLVRGNHESRQITQVYGFYEECLNKYGSTTVWKYCCQVFDFLTLAAIIDGKI
                     LCVHGGLSPEIRMLDQIRVLSRAQEVPHEGGFSDLLWSDPDNVEAWQVSPRGAGWLFG
                     SKVAREFNHVNGLNLIARAHQLVMEGFKYHFPEKDVVTVWSAPNYCYRCGNVASVMKV
                     DEDLEPTFKIFSAVPDDYIRESTANHNNQRAGYFL"
     CDS             3582. .4103
                     /standard_name="D2699"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96443.1"
                     /db_xref="GI:1279670"
                     /db_xref="SPTREMBL:Q12408"
                     /translation="MTHSLKALFALLFLYTAAVNAGVIGIFNALPPPNTKPINGESPL
                     YQCDILDKQLVEIKEVNLDPNPPVRGENLTISANGEVFETIEEGAYIDVEVRLGYIRL
                     LSQTFDLCETLEDNDIEGLSCPIEPGEYNIKKIVEIPGEVPPGKYVVVARAYTEKDDL
                     ITCLTGEVIFPPR"
     gene            complement(5030. .5950)
                     /gene="FAD1"
                     /db_xref="SGD:S0002203"
     CDS             complement(5030. .5950)
                     /gene="FAD1"
                     /standard_name="D2702"
                     /note="Acc.no. U12331"
                     /codon_start=1
                     /db_xref="SGD:S0002203"
                     /product="FAD synthetase"
                     /protein_id="CAA96444.1"
                     /db_xref="GI:1279671"
                     /db_xref="SWISS-PROT:P38913"
                     /translation="MQLSKAAEMCYEITNSYLHIDQKSQIIASTQEAIRLTRKYLLSE
                     IFVRWSPLNGEISFSYNGGKDCQVLLLLYLSCLWEYFFIKAQNSQFDFEFQSFPMQRL
                     PTVFIDQEETFPTLENFVLETSERYCLSLYESQRQSGASVNMADAFRDFIKIYPETEA
                     IVIGIRHTDPFGEALKPIQRTDSNWPDFMRLQPLLHWDLTNIWSFLLYSNEPICGLYG
                     KGFTSIGGINNSLPNPHLRKDSNNPALHFEWEIIHAFGKDAEGERSSAINTSPISVVD
                     KERFSKYHDNYYPGWYLVDDTLERAGRIKN"
     gene            complement(6581. .7630)
                     /gene="NAM1"
                     /db_xref="SGD:S0002202"
     CDS             complement(6581. .7630)
                     /gene="NAM1"
                     /standard_name="D2705"
                     /codon_start=1
                     /db_xref="SGD:S0002202"
                     /protein_id="CAA96445.1"
                     /db_xref="GI:1279672"
                     /db_xref="SWISS-PROT:P10849"
                     /translation="MIRTSSILKNCNYRYIHCIHRCLLNEANLKDRKTHNVERVSNEK
                     TFEQALEEERKVFGELFEAGARVENMRHTNASKIIDKYYNGLQDNSEGTSVKKEKIVF
                     NHSQRAQRKLPNKDHEFLKETAGNDYVYERAEPSAISTKTISEQTRTLLEKIFDEDNS
                     INKSNRELLNLNLRKGSGMEALRQPVAHSNVKFSEEVMQEIGNKIRYQTTLDQVLEPH
                     IDYLREAVKSDYDLLRYLKQSLDIYKKRNKDLELKMNAESSNIFEDIRSACINKPAEL
                     PKPLAMTLPYIIVKSLRLGDFDFPADRKYTLISYVYNECKNNMDASLYLTICNVDFYN
                     LLVQLLWETSRRFAT"
     gene            complement(8021. .8812)
                     /gene="RNA11"
                     /db_xref="SGD:S0002201"
     CDS             complement(8021. .8812)
                     /gene="RNA11"
                     /standard_name="D2711"
                     /note="Protein sequence is in conflict with the conceptual
                     translation; Acc.no. X01419"
                     /codon_start=1
                     /db_xref="SGD:S0002201"
                     /product="RNA11"
                     /protein_id="CAA96446.1"
                     /db_xref="GI:1279673"
                     /db_xref="SWISS-PROT:Q07350"
                     /translation="LEGVGSKKGGGGIASESQFNLQRRKEVESLLSKGENVPYTFQDE
                     KDDQVRSNPYIYKNHSGKLVCKLCNTMHMSWSSVERHLGGKKHGLNVLRRGISIEKSS
                     LGREGQTTHDFRQQQKIIEAKQSLKNNGTIPVCKIATVKNPKNGSVGLAIQVNYSSEV
                     KENSVDSDDKAKVPPLIRIVSGLELSDTKQKGKKFLVIAYEPFENIAIELPPNEILFS
                     ENNDMDNNNDGVDELNKKCTFWDAISKLYYVQFFFKQAEQEQADV"
     gene            complement(9098. .10786)
                     /gene="SIR2"
                     /db_xref="SGD:S0002200"
     CDS             complement(9098. .10786)
                     /gene="SIR2"
                     /standard_name="D2714"
                     /note="Acc.no. X01419"
                     /codon_start=1
                     /db_xref="SGD:S0002200"
                     /product="SIR2"
                     /protein_id="CAA96447.1"
                     /db_xref="GI:1279674"
                     /db_xref="SWISS-PROT:P06700"
                     /translation="MTIPHMKYAVSKTSENKVSNTVSPTQDKDAIRKQPDDIINNDEP
                     SHKKIKVAQPDSLRETNTTDPLGHTKAALGEVASMELKPTNDMDPLAVSAASVVSMSN
                     DVLKPETPKGPIIISKNPSNGIFYGPSFTKRESLNARMFLKYYGAHKFLDTYLPEDLN
                     SLYIYYLIKLLGFEVKDQALIGTINSIVHINSQERVQDLGSAISVTNVEDPLAKKQTV
                     RLIKDLQRAINKVLCTRLRLSNFFTIDHFIQKLHTARKILVLTGAGVSTSLGIPDFRS
                     SEGFYSKIKHLGLDDPQDVFNYNIFMHDPSVFYNIANMVLPPEKIYSPLHSFIKMLQM
                     KGKLLRNYTQNIDNLESYAGISTDKLVQCHGSFATATCVTCHWNLPGERIFNKIRNLE
                     LPLCPYCYKKRREYFPEGYNNKVGVAASQGSMSERPPYILNSYGVLKPDITFFGEALP
                     NKFHKSIREDILECDLLICIGTSLKVAPVSEIVNMVPSHVPQVLINRDPVKHAEFDLS
                     LLGYCDDIAAMVAQKCGWTIPHKKWNDLKNKNFKCQEKDKGVYVVYTSDEHPKTL"
     CDS             10443. .10796
                     /standard_name="D2717"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96448.1"
                     /db_xref="GI:1279675"
                     /db_xref="SPTREMBL:Q12352"
                     /translation="MIGPLGVSGFKTSLDIDTTEAADTAKGSMSLVGLSSIDATSPSA
                     ALVCPSGSVVLVSLKESGCATFIFLCEGSSLFIMSSGCFLIASLSCVGLTVFETLFSL
                     VFDTAYFICGMVIQL"
     gene            complement(11215. .13779)
                     /gene="NAT1"
                     /db_xref="SGD:S0002198"
     CDS             complement(11215. .13779)
```

```
                     /gene="NAT1"
                     /standard_name="D2720"
                     /note="Acc.no. X15135"
                     /codon_start=1
                     /db_xref="SGD:S0002198"
                     /product="N-terminal acetyltransferase"
                     /protein_id="CAA96449.1"
                     /db_xref="GI:1279676"
                     /db_xref="SWISS-PROT:P12945"
                     /translation="MSRKRSTKPKPAAKIALKKENDQFLEALKLYEGKQYKKSLKLLD
                     AILKKDGSHVDSLALKGLDLYSVGEKDDAASYVANAIRKIEGASASPICCHVLGIYMR
                     NTKEYKESIKWFTAALNNGSTNKQIYRDLATLQSQIGDFKNALVSRKKYWEAFLGYRA
                     NWTSLAVAQDVNGERQQAINTLSQFEKLAEGKISDSEKYEHSECLMYKNDIMYKAASD
                     NQDKLQNVLKHLNDIEPCVFDKFGLLERKATIYMKLGQLKDASIVYRTLIKRNPDNFK
                     YYKLLEVSLGIQGDNKLKKALYGKLEQFYPRCEPPKFIPLTFLQDKEELSKKLREYVL
                     PQLERGVPATFSNVKPLYQRRKSKVSPLLEKIVLDYLSGLDPTQDPIPFIWTNYYLSQ
                     HFLFLKDFPKAQEYIDAALDHTPTLVEFYILKARILKHLGLMDTAAGILEEGRQLDLQ
                     DRFINCKTVKYFLRANNIDKAVEVASLFTKNDDSVNGIKDLHLVEASWFIVEQAEAYY
                     RLYLDRKKKLDDLASLKKEVESDKSEQIANDIKENQWLVRKYKGLALKRFNAIPKFYK
                     QFEDDQLDFHSYCMRKGTPRAYLEMLEWGKALYTKPMYVRAMKEASKLYFQMHDDRLK
                     RKSDSLDENSDEIQNNGQNSSSQKKKAKKEAAAMNKRKETEAKSVAAYPSDQDNDVFG
                     EKLIETSTPMEDFATEFYNNYSMQVREDERDYILDFEFNYRIGKLALCFASLNKFAKR
                     FGTTSGLFGSMAIVLLHATRNDTPFDPILKKVVTKSLEKEYSENFPLNEISNNSFDWL
                     NFYQEKFGKNDINGLLFLYRYRDDVPIGSSNLKEMIISSLSPLEPHSQNEILQYYL"
     CDS             complement(14327. .14674)
                     /standard_name="D2723"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA96450.1"
                     /db_xref="GI:1279677"
                     /db_xref="SPTREMBL:Q12459"
                     /translation="MTSPASTSTISNVQSTASVMNHSIEDNISAAASLESVSGTSTKD
```

Query Match 8.4%; Score 18; DB 7; Length 36687;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db  13036 TGCTTTTCTCTCTAATAA 13053
```

```
RESULT   9
SPBC660
LOCUS       SPBC660     43325 bp    DNA             PLN       31-JAN-2000
DEFINITION  S.pombe chromosome II cosmid c660.
ACCESSION   AL034563
VERSION     AL034563.1  GI:4049499
KEYWORDS    6-phosphogluconate dehydrogenase decarboxylating; cell wall
            protein; class v pyridoxal phosphate dependent aminotransferase;
            elongation factor g; elongation factor Tu family; fbp1;
            fructose-1,6-bisphosphatase; G beta repeat; glycine-rich protein;
            low-complexity gene-free region; mik1; mitosis inhibitor protein
            kinase mik1; myb like dna-binding protein; neutral trehalase; ntp1;
            polyA-binding protein; ras11; replication factor-a protein 1;
            ribonucleoprotein; RNA recognition; RNA3' Cleavage factor Ib; rpa1;
            ssb1; transcription initiation factor iif beta subunit; WD domain;
            yeast CF Ib.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
            Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE   1  (bases 1 to 43325)
  AUTHORS   Lyne,M., Rajandream,M.A., Barrell,B.G. and Rieger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-1998) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
            and Biotechnologische und molekularbiologische Forschung,
            Angelhofweg 39, D-69259 Wilhelmsfeld, Germany
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            During 1995 to 1996 about 66% of S. pombe chromosome 1 was
            sequenced by the Sanger Centre.  The sequencing of the S. pombe
            genome is now being continued with funding from The European
            Commission. Fourteen European sequencing laboratories, including
            the Sanger Centre, are participating in the project.
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
            pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c660 is overlapped at the 3' end by cosmid 1198 (contained
            in EMBL entry SP33010 accession number U33010).
FEATURES             Location/Qualifiers
     source          1. .43325
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="II"
                     /clone="cosmid c660"
                     /map="IIL"
     gene            complement(1. .1482)
                     /gene="SPBC660.01c"
     CDS             complement(join(1. .86,138. .464,506. .1482))
                     /partial
                     /gene="SPBC660.01c"
                     /note="SPBC660.01c, SIMILARITY:Schizosaccharomyces pombe,
                     CAB52717, putative myb-like dna-binding protein., (496
                     aa), fasta scores: opt: 478, E():4.7e-23, (30.6% identity
                     in 350 aa)"
                     /codon_start=1
                     /label=SPBC660.01c
                     /product="putative myb like dna-binding protein"
                     /protein_id="CAA22521.1"
                     /db_xref="GI:4049500"
                     /db_xref="SPTREMBL:O94422"
                     /translation="MDTSVLNPELQIHGFIGVDSLQSSRKRKNDFDDFPLNKGLKTNN
                     NDYSGSIEPKFSPALSIKEDGKNDRNFEALMSLQAQDSNLSEQNTSIHLDALASSSIA
                     LGNDNVDSSAFLSKVNKGVNAMRNSTSNQTNDSILISPSEITNMDPFLKGSARWTAEH
                     WDYLERRMQNFCQTYSLDHTQVADSLHEKRLHGPLSSLVKLLVQEMPSFTRRTILRHL
                     RALYNIPGYEKYSRKNSSGRGDFGVQETAIISQEVHNFIMDQGWSEYQFCNQIWAGKC
                     PKTIRMFYSNLYKKLSHRDAKSIYHHVRRAYNPFEDRCVWSKEEDEELRKNVVEHGKC
                     WTKIGRKMARMPNDCRDRWRDVVRFGDKLKRNAWSLEEETQLLQIVAELRNREDLSSD
                     INWTLVAQMLGTRTRLQCRYKFQQLTKAASKFELQENVWLLERIYDSLLNNGGKIHWE
                     NIVKEANGRWTRD"
     misc_feature    1. .7073
                     /note="nominal overlap with cosmids SP33010, EM:U33010 S.
                     pombe chromosome 2"
     misc_feature    complement(87. .104)
                     /gene="SPBC660.01c"
                     /note="ctaatattttaatttaag, splice branch and acceptor"
     misc_feature    complement(132. .137)
                     /gene="SPBC660.01c"
                     /note="gtaagt, splice donor sequence"
     misc_feature    complement(join(383. .464,506. .537))
                     /gene="SPBC660.01c"
                     /note="Match to PF00249 myb_DNA-binding, Myb-like
                     DNA-binding domain Score 30.86"
     misc_feature    complement(465. .475)
                     /gene="SPBC660.01c"
                     /note="ttaacgtttag, splice branch and acceptor"
```

misc_feature complement(500. .505)
/gene="SPBC660.01c"
/note="gtatgt, splice donor sequence"
gene 2618. .3929
/gene="SPBC660.02"
CDS join(2618. .2738,2785. .3929)
/gene="SPBC660.02"
/note="SPBC660.02, len:421, SIMILARITY:Schizosaccharomyces pombe, O13286, srw1., (556 aa), fasta scores: opt: 1364, E():0, (50.6% identity in 385 aa)"
/codon_start=1
/label=SPBC660.02
/product="WD domain; G beta repeat protein"
/protein_id="CAA22522.1"
/db_xref="GI:4049501"
/db_xref="SPTREMBL:O94423"
/translation="MGDRFIPIRNVSNEFNFSFQSFKECVLSHGSNLRRKTSGTIQRQ
FMELLSMELFGSQASRSRAFYYGEDKRKIEKKMLDTPDRKSYSLSPISPQSQDMLRQP
QKPKRAFPKTPYKILDAPYLKNDFYLNLLDWGQSNVLAVGLASSIYLWSAASGKVVQL
HDFGATNHVTSVLWTGKGTQLAVGTDSGVIYIWDIESTKSVRSLKGHSERVAALAWND
NTLTSGGKDEVILHHDLRAPGCCAEMMKVHEQEICGLQWDRSLGQLASGGNDNNLFVW
DYRSSRPLHKFEEHTAAVKAIGWSPHQRGILASGGGTIDRCLTIHNTLTGRLQNKLDT
GSQVCNMAWSKTSNEIVTTHGFAKNQVSLWKYPSLKNIANLTAHTNRVLYLSMSPDGQ
SIVTGAGDETLRFWKLFNKKPKEESTLIR"
misc_feature 2739. .2744
/gene="SPBC660.02"
/note="gtatgt, splice donor sequence"
misc_feature 2772. .2784
/gene="SPBC660.02"
/note="ctaacgacagcag, splice branch and acceptor"
misc_feature 3258. .3353
/gene="SPBC660.02"
/note="Match to PF00400 WD40, WD domain, G-beta repeat Score 22.93"
misc_feature 3378. .3494
/gene="SPBC660.02"
/note="Match to PF00400 WD40, WD domain, G-beta repeat Score 31.68"
misc_feature 3771. .3884
/gene="SPBC660.02"
/note="Match to PF00400 WD40, WD domain, G-beta repeat Score 20.22"
gene complement(3930. .5015)
/gene="SPBC660.03c"
CDS complement(join(3930. .4184,4288. .4431,4491. .5015))
/gene="SPBC660.03c"
/note="SPBC660.03c, len:307, SIMILARITY:Saccharomyces cerev isiae, YGR005C, T2FB_YEAST, transcription initiation factor iif, beta subunit, (400 aa), fasta scores: opt: 461, E():2 .9e-32, (31.0% identity in 368 aa); SPBC660.03c, len:307, SIMILARITY:Saccharomyces cerevisiae, T2FB_YEAST, transcription initiation factor iif, beta subunit, (400 aa), fasta scores: opt: 461, E():1.4e-22, (31.0% identity in 368 aa)"
/codon_start=1
/label=SPBC660.03c
/product="transcription initiation factor iif, beta subunit"
/protein_id="CAA22523.1"
/db_xref="GI:4049502"
/db_xref="SPTREMBL:O94424"
/translation="MSEEKPTVRTEEDDRYEDDAGDLDLGQIGSRVWLVKIPKFLMDK
WNSIPEDDAANLGCVRVKNDEIQLLLQNSPENADVPKIYNLRVMNKFVRNSYVFRESE
TSSSMKSTALVGTVAHECNVSPVINDDYRRVMQKRALAASAPKRKVQMIDDRGGSLLA
PGTLGSRSRSTTSFIRNVKPRTGEGLKNSRIPRNELLDILFKCFEDYEYWTLKGLREY
VKQPEVYLKEVLDSIAILNKRGPYALKYSLKPEYKGTMDAASVELRNQQASQSESSSI
DHTGKNTSPDNPGTNAEEDEDDDGVEMIDVV"
misc_feature complement(4185. .4202)
/gene="SPBC660.03c"
/note="ctaatgaattcatattag, splice branch and acceptor"
misc_feature complement(4282. .4287)
/gene="SPBC660.03c"
/note="gtatgt, splice donor sequence"
misc_feature complement(4432. .4447)
/gene="SPBC660.03c"
/note="ctaacttttttttcag, splice branch and acceptor"
misc_feature complement(4485. .4490)
/gene="SPBC660.03c"
/note="gtatgt, splice donor sequence"
gene complement(6039. .7082)
/gene="SPBC660.04c"
CDS complement(6039. .7082)
/gene="SPBC660.04c"
/note="SPBC660.04c, len:347; SPBC660.04c, len:347"
/codon_start=1
/label=fbp1
/product="fructose-1,6-bisphosphatase"
/protein_id="CAA22524.1"
/db_xref="GI:4049503"
/translation="MKKDLDEIDTDIVTLSSFILQEQRRYNQKHKNEEGKPCIIQEAS
GELSLLLNSLQFSFKFIANTIRKAELVNLIGLSGIVNSTGDEQKKLDKICNDIFITAM
KSNGCCKLIVSEEEEDLIVVDSNGSYAVTCDPIDGSSNIDAGVSVGTIFGIYKLRPGS
QGDISDVLRPGKEMVAAGYTMYGASAHLLLTTGHRVNGFTLDTDIGEFILTHRNMKMP
LQHSIYSINEGYTAFWDEKIARFIAHLKESTPDKKPYSARYIGSMVADMHRTILYGGL
FAYPCSKGNNGKLRLLYECFPMAFLVEQAGGIAVNDKGDRILDLVPKTLHGKSSIWLG
SKHEVEEYINFIK"
misc_feature complement(6066. .6833)
/gene="SPBC660.04c"

Query Match 8.4%; Score 18; DB 8; Length 43325;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 AAGAAAACATCTACTTTG 201
||||||||||||||||||
Db 22202 AAGAAAACATCTACTTTG 22219

RESULT 10
AC022747/c
LOCUS AC022747 83536 bp DNA HTG 06-FEB-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-131K9 map 4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC022747
VERSION AC022747.1 GI:6987626
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 83536)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-131K9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 83536)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:

```
            128533    130201 contig of   1669 bp in length
            130702    131963 contig of   1262 bp in length
            132464    134276 contig of   1813 bp in length
            134777    135777 contig of   1001 bp in length
            136278    137928 contig of   1651 bp in length
            138429    139740 contig of   1312 bp in length
          Sequence updated (16-Feb-2000).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 25 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    15833: contig of 15833 bp in length
          *    15834 16333: gap of      500 bp
          *    16334    34076: contig of 17743 bp in length
          *    34077 34581: gap of      505 bp
          *    34582    46531: contig of 11950 bp in length
          *    46532 47034: gap of      503 bp
          *    47035    56602: contig of 9568 bp in length
          *    56603 57107: gap of      505 bp
          *    57108    65754: contig of 8647 bp in length
          *    65755 66255: gap of      501 bp
          *    66256    71355: contig of 5100 bp in length
          *    71356 71866: gap of      511 bp
          *    71867    77672: contig of 5806 bp in length
          *    77673 78174: gap of      502 bp
          *    78175    81813: contig of 3639 bp in length
          *    81814 82313: gap of      500 bp
          *    82314    87384: contig of 5071 bp in length
          *    87385 87886: gap of      502 bp
          *    87887    92260: contig of 4374 bp in length
          *    92261 92762: gap of      502 bp
          *    92763    97056: contig of 4294 bp in length
          *    97057 97566: gap of      510 bp
          *    97567   102247: contig of 4681 bp in length
          *   102248 102760: gap of      513 bp
          *   102761   106861: contig of 4101 bp in length
          *   106862 107367: gap of      506 bp
          *   107368   109928: contig of 2561 bp in length
          *   109929 110444: gap of      516 bp
          *   110445   113790: contig of 3346 bp in length
          *   113791 114290: gap of      500 bp
          *   114291   117299: contig of 3009 bp in length
          *   117300 117801: gap of      502 bp
          *   117802   121336: contig of 3535 bp in length
          *   121337 121836: gap of      500 bp
          *   121837   125424: contig of 3588 bp in length
          *   125425 125924: gap of      500 bp
          *   125925   128031: contig of 2107 bp in length
          *   128032 128536: gap of      505 bp
          *   128537   130196: contig of 1660 bp in length
          *   130197 130709: gap of      513 bp
          *   130710   131962: contig of 1253 bp in length
          *   131963 132465: gap of      503 bp
          *   132466   134274: contig of 1809 bp in length
          *   134275 134778: gap of      504 bp
          *   134779   135777: contig of 999 bp in length
          *   135778 136277: gap of      500 bp
          *   136278   137927: contig of 1650 bp in length
          *   137928 138429: gap of      502 bp
          *   138430   139740: contig of 1311 bp in length.
FEATURES             Location/Qualifiers
     source          1. .139740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="CMB9-21K9"
                     /map="11q22"
BASE COUNT    40055 a  21062 c  21342 g  43261 t  14020 others
ORIGIN
```

```
  Query Match             8.4%;  Score 18;  DB 31;  Length 139740;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    168 TGCTTTTCTCTCTAATAA 185
          ||||||||||||||||||
Db  94300 TGCTTTTCTCTCTAATAA 94283


RESULT  15
AC022414
LOCUS       AC022414   141107 bp    DNA             HTG       21-MAR-2000
DEFINITION  Homo sapiens chromosome 5 clone CTC-316M18, WORKING DRAFT SEQUENCE,
            30 unordered pieces.
ACCESSION   AC022414
VERSION     AC022414.2  GI:7272110
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141107)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141107)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 21, 2000 this sequence version replaced gi:6862782.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------Summary Statistics
            Consensus quality: 118702 bases at least Q40
            Consensus quality: 130365 bases at least Q30
            Consensus quality: 133293 bases at least Q20
            Estimated insert size: 141107; sum-of-contigs estimation
            Estimated insert size: 149000; pulse field gel estimation
            Quality coverage:  3.58x in Q20 bases; pulse field gel estimation
            Quality coverage:  3.78x in Q20 bases; sum-of-contigs estimation
            -------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1133: contig of 1133 bp in length
            *                     gap of unknown length
            *     1134     2199: contig of 1066 bp in length
            *                     gap of unknown length
            *     2200     3248: contig of 1049 bp in length
            *                     gap of unknown length
            *     3249     4300: contig of 1052 bp in length
            *                     gap of unknown length
            *     4301     5524: contig of 1224 bp in length
            *                     gap of unknown length
            *     5525     6528: contig of 1004 bp in length
            *                     gap of unknown length
            *     6529     7824: contig of 1296 bp in length
            *                     gap of unknown length
            *     7825     8934: contig of 1110 bp in length
            *                     gap of unknown length
            *     8935    10145: contig of 1211 bp in length
            *                     gap of unknown length
            *    10146    11196: contig of 1051 bp in length
```

```
            *                    gap of unknown length
            *    11197    12478: contig of 1282 bp in length
            *                    gap of unknown length
            *    12479    14044: contig of 1566 bp in length
            *                    gap of unknown length
            *    14045    16178: contig of 2134 bp in length
            *                    gap of unknown length
            *    16179    17561: contig of 1383 bp in length
            *                    gap of unknown length
            *    17562    19355: contig of 1794 bp in length
            *                    gap of unknown length
            *    19356    21747: contig of 2392 bp in length
            *                    gap of unknown length
            *    21748    24712: contig of 2965 bp in length
            *                    gap of unknown length
            *    24713    27513: contig of 2801 bp in length
            *                    gap of unknown length
            *    27514    32242: contig of 4729 bp in length
            *                    gap of unknown length
            *    32243    38203: contig of 5961 bp in length
            *                    gap of unknown length
            *    38204    41038: contig of 2835 bp in length
            *                    gap of unknown length
            *    41039    44977: contig of 3939 bp in length
            *                    gap of unknown length
            *    44978    52346: contig of 7369 bp in length
            *                    gap of unknown length
            *    52347    59597: contig of 7251 bp in length
            *                    gap of unknown length
            *    59598    66217: contig of 6620 bp in length
            *                    gap of unknown length
            *    66218    74249: contig of 8032 bp in length
            *                    gap of unknown length
            *    74250    86682: contig of 12433 bp in length
            *                    gap of unknown length
            *    86683    99513: contig of 12831 bp in length
            *                    gap of unknown length
            *    99514   120052: contig of 20539 bp in length
            *                    gap of unknown length
            *   120053   141107: contig of 21055 bp in length.
FEATURES             Location/Qualifiers
     source          1. .141107
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-316M18"
BASE COUNT    42267 a  28713 c  29170 g  40884 t     73 others
ORIGIN
```

```
  Query Match             8.4%;  Score 18;  DB 67;  Length 141107;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indles    0;. Gaps    0;

Qy      93 CTTTGCAGAGAGAACAGC 110
           ||||||||||||||||||
Db   42471 CTTTGCAGAGAGAACAGC 42488


Search completed: October  3, 2000, 12:52:26
Job time: 8948 sec
```

```
repeat_region   1410. .1587
                /note="MLT1D repeat: matches 139. .323 of consensus"
repeat_region   1575. .1858
                /note="MLT1E repeat: matches 265. .568 of consensus"
repeat_region   1876. .2745
                /note="L1MB5 repeat: matches 921. .1 of consensus"
repeat_region   2597. .2793
                /note="L1 repeat: matches 5390. .5194 of consensus"
repeat_region   2854. .3352
                /note="L1ME1 repeat: matches 687. .170 of consensus"
repeat_region   2971. .3352
                /note="L1ME2 repeat: matches 570. .170 of consensus"
repeat_region   3327. .3559
                /note="MER20 repeat: matches 1. .218 of consensus"
repeat_region   3580. .3873
                /note="AluSc repeat: matches 1. .299 of consensus"
repeat_region   3877. .4060
                /note="L1MC2 repeat: matches 193. .10 of consensus"
repeat_region   6468. .6575
                /note="MIR2 repeat: matches 22. .145 of consensus"
repeat_region   7024. .7238
                /note="AluY repeat: matches 84. .299 of consensus;
                incomplete repeat"
repeat_region   7834. .8126
                /note="AluSc repeat: matches 2. .295 of consensus"
repeat_region   10054. .10223
                /note="MIR repeat: matches 89. .262 of consensus"
misc_feature    complement(10749. .11106)
                /note="match: STS G05144"
repeat_region   11725. .12081
                /note="THE1C repeat: matches 371. .1 of consensus"
repeat_region   12196. .12497
                /note="AluSq repeat: matches 303. .1 of consensus"
repeat_region   13253. .13367
                /note="MIR2 repeat: matches 4. .127 of consensus"
repeat_region   13398. .15749
                /note="L1 repeat: matches 2577. .4940 of consensus"
repeat_region   16076. .16554
                /note="L1 repeat: matches 4910. .5390 of consensus"
repeat_region   16409. .17294
                /note="L1PA2 repeat: matches 1. .891 of consensus"
repeat_region   17808. .19825
                /note="L1 repeat: matches 3338. .5390 of consensus"
repeat_region   19683. .20571
                /note="L1PA6 repeat: matches 1. .890 of consensus"
repeat_region   22732. .22827
                /note="L1MB4 repeat: matches 797. .902 of consensus"
repeat_region   22748. .22851
                /note="L1MB5 repeat: matches 812. .922 of consensus"
prim_transcript <25068. .25543
                /note="match: H61071 H69565"
repeat_region   25302. .25381
                /note="MIR repeat: matches 49. .131 of consensus"
repeat_region   25527. .25703
                /note="MER45 repeat: matches 1. .178 of consensus"
repeat_region   25932. .26062
                /note="MLT1F repeat: matches 18. .138 of consensus"
repeat_region   26055. .26264
                /note="L1 repeat: matches 2971. .2759 of consensus"
repeat_region   26266. .26378
                /note="L1MA7 repeat: matches 1017. .895 of consensus"
prim_transcript 27046. .>27579
                /note="match: multiple ESTs; match: R71060 H69028 R82016
                R10102 R82280 R82281; match: R82066 H69792 H79486 AA506861
                T78091; match: W03874 H74282 D85329"
CDS             complement(join(27431. .27577,28562. .28744,31530. .31681,
                33638. .33782,36315. .36470,36919. .37022,37955. .38026,
                39019. .39135,41033. .41212,42726. .42936,43904. .44140,
                49624. .49798,53412. .56232,57414. .57626,59560. .59710,
                62919. .63133,63758. .63857,65675. .65852,68300. .68465,
                69764. .69985,72271. .72414,73672. .73884,85339. .85461,
                95549. .95640,99347. .99504))
                /codon_start=1
                /product="factor V"
                /protein_id="CAB16748.1"
                /db_xref="GI:2769647"
                /db_xref="SPTREMBL:O43737"
                /translation="MFPGCPRLWVLVVLGTSWVGWGSQGTEAAQLRQFYVAAQGISWS
                YRPEPTNSSLNLSVTSFKKIVYREYEPYFKKEKPQSTISGLLGPTLYAEVGDIIKVHF
                KNKADKPLSIHPQGIRYSKLSEGASYLDHTFPAEKMDDAVAPGREYTYEWSISEDSGP
                THDDPPCLTHIYYSHENLIEDFNSGLIGPLLICKKGTLTEGGTQKTFDKQIVLLFAVF
                DESKSWSQSSSLMYTVNGYVNGTMPDITVCAHDHISWHLLGMSSGPELFSIHFNGQVL
                EQNHHKVSAITLVSATSTTANMTVGPEGKWIISSLTPKHLQAGMQAYIDIKNCPKKTR
                NLKKITREQRRHMKRWEYFIAAEEVIWDYAPVIPANMDKKYRSQHLDNFSNQIGKHYK
                KVMYTQYEDESFTKHTVNPNMKEDGILGPIIRAQVRDTLKIVFKNMASRPYSIYPHGV
                TFSPYEDEVNSSFTSGRNNTMIRAVQPGETYTYKWNILEFDEPTENDAQCLTRPYYSD
                VDIMRDIASGLIGLLLICKSRSLDRQGIQRAADIEQQAVFAVFDENKSWYLEDNINKF
                CENPDEVKRDDPKFYESNIMSTINGYVPESITTLGFCFDDTVQWHFCSVGTQNEILTI
                HFTGHSFIYGKRHEDTLTLFPMRGESVTVTMDNVGTWMLTSMNSSPRSKKLRLKFRDV
                KCIPDDDEDSYEIFEPPESTVMATRKMHDRLEPEDEESDADYDYQNRLAAALGIRSFR
                NSSLNQEEEEFNLTALALENGTEFVSSNTDIIVGSNYSSPSNISKFTVNNLAEPQKAP
                SHQQATTAGSPLRHLIGKNSVLNSSTAEHSSPYSEDPIEDPLQPDVTGIRLLSLGAGE
                FKSQEHAKHKGPKVERDQAAKHRFSWMKLLAHKVGRHLSQDTGSPSGMRPWEDLPSQD
                TGSPSRMRPWKDPPSDLLLLKQSNSSKILVGRWHLASEKGSYEIIQDTDEDTAVNNWL
                ISPQNASRAWGESTPLANKPGKQSGHPKFPRVRHKSLQVRQDGGKSRLKKSQFLIKTR
                KKKKEKHTHHAPLSPRTFHPLRSEAYNTFSERRLKHSLVLHKSNETSLPTDLNQTLPS
                MDFGWIASLPDHNQNSSNDTGQASCPPGLYQTVPPEEHYQTFPIQDPDQMHSTSDPSH
                RSSSPELSEMLEYDRSHKSFPTDISQMSPSSEHEVWQTVISPDLSQVTLSPELSQTNL
                SPDLSHTTLSPELIQRNLSPALGQMPISPDLSHTTLSPDLSHTTLSLDLSQTNLSPEL
                SQTNLSPALGQMPLSPDLSHTTLSLDFSQTNLSPELSHMTLSPELSQTNLSPALGQMP
                ISPDLSHTTLSLDFSQTNLSPELSQTNLSPALGQMPLSPDPSHTTLSLDLSQTNLSPE
                LSQTNLSPDLSEMPLFADLSQIPLTPDLDQMTLSPDLGETDLSPNFGQMSLSPDLSQV
                TLSPDISDTTLLPDLSQISPPPDLDQIFYPSESSQSLLLQEFNESFPYPDLGQMPSPS
                SPTLNDTFLSKEFNPLVIVGLSKDGTDYIEIIPKEEVQSSEDDYAEIDYVPYDDPYKT
                DVRTNINSSRDPDNIAAWYLRSNNGNRRNYYIAAEEISWDYSEFVQRETDIEDSDDIP
                EDTTYKKVVFRKYLDSTFTKRDPRGEYEEHLGILGPIIRAEVDDVIQVRFKNLASRPY
                SLHAHGLSYEKSSEGKTYEDDSPEWFKEDNAVQPNSSYTYVWHATERSGPESPGSACR
                AWAYYSAVNPEKDIHSGLIGPLLICQKGILHKDSNMPMDMREFVLLFMTFDEKKSWYY
                EKKSRSSWRLTSSEMKKSHEFHAINGMIYSLPGLKMYEQEWVRLHLLNIGGSQDIHVV
                HFHGQTLLENGNKQHQLGVWPLLPGSFKTLEMKASKPGWWLLNTEVGENQRAGMQTPF
                LIMDRDCRMPMGLSTGIISDSQIKASEFLGYWEPRLARLNNGGSYNAWSVEKLAAEFA
                SKPWIQVDMQKEVIITGIQTQGAKHYLKSCYTTEFYVAYSSNQINWQIFKGNSTRNVM
                YFNGNSDASTIKENQFDPPIVARYIRISPTRAYNRPTLRLELQGCEVNGCSTPLGMEN
                GKIENKQITASSFKKSWWGDYWEPFRARLNAQGRVNAWQAKANNNKQWLEIDLLKIKK
                ITAIITQGCKSLSSEMYVKSYTIHYSEQGVEWKPYRLKSSMVDKIFEGNTNTKGHVKN
                FFNPPIISRFIRVIPKTWNQSIALRLELFGCDIY"
repeat_region   27629. .27788
                /note="AluJo repeat: matches 302. .140 of consensus;
                incomplete repeat"
repeat_region   28002. .28176
                /note="MIR repeat: matches 75. .259 of consensus"
repeat_region   29041. .29084
                /note="MIR2 repeat: matches 103. .146 of consensus"
repeat_region   31903. .31988
                /note="MIR repeat: matches 104. .17 of consensus"
repeat_region   34738. .35036
                /note="AluSc repeat: matches 299. .1 of consensus"
repeat_region   35051. .35542
                /note="L1 repeat: matches 3215. .3729 of consensus"
repeat_region   36628. .36808
                /note="MIR repeat: matches 40. .235 of consensus"
repeat_region   38333. .38451
                /note="AluJo repeat: matches 1. .134 of consensus;
                incomplete repeat"
repeat_region   40215. .40272
                /note="MADE1 repeat: matches 1. .53 of consensus"
repeat_region   43274. .43327
                /note="MIR2 repeat: matches 146. .93 of consensus"
repeat_region   43456. .43548
                /note="L1MA8 repeat: matches 1038. .944 of consensus"
repeat_region   44951. .46774
                /note="L1 repeat: matches 3613. .5390 of consensus"
repeat_region   46627. .46980
                /note="L1MA7 repeat: matches 2. .364 of consensus"
repeat_region   46983. .47446
```

```
  Query Match             8.4%;  Score 18;  DB 10;  Length 106571;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    180 TAATAAGAAAACATCTAC 197
          ||||||||||||||||||
Db   1250 TAATAAGAAAACATCTAC 1233


RESULT 13
AC010429
LOCUS       AC010429   133783 bp    DNA             HTG       05-APR-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2199L14, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC010429
VERSION     AC010429.3  GI:7417548
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133783)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133783)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:7212886.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------Summary Statistics
            Consensus quality: 132291 bases at least Q40
            Consensus quality: 133603 bases at least Q30
            Consensus quality: 133733 bases at least Q20
            Estimated insert size: 133783; sum-of-contigs estimation
            Estimated insert size: 233000; pulse field gel estimation
            Quality coverage:  4.87x in Q20 bases; pulse field gel estimation
            Quality coverage:  8.48x in Q20 bases; sum-of-contigs estimation
            -------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9662: contig of 9662 bp in length
            *                   gap of unknown length
            *    9663   133783: contig of 124121 bp in length.
FEATURES            Location/Qualifiers
     source         1. .133783
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTD-2199L14"
BASE COUNT    43496 a  25289 c  24707 g  40291 t
ORIGIN


  Query Match             8.4%;  Score 18;  DB 72;  Length 133783;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    187 AAAACATCTACTTTGAAA 204
          ||||||||||||||||||
Db  62087 AAAACATCTACTTTGAAA 62104


RESULT 14
AP000817/c
LOCUS       AP000817   139740 bp    DNA             HTG       18-FEB-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-21K9 map 11q22, WORKING DRAFT
            SEQUENCE, 25 unordered pieces.
ACCESSION   AP000817
VERSION     AP000817.2  GI:7007459
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:CMB9-21K9.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139740)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 139,740 genomic DNA of 11q22
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 139740)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Feb 19, 2000 this sequence version replaced gi:6997652.
            ---------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.RIKEN.go.jp/
                Contact: hattori@gsc.RIKEN.go.jp
            ---------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-21K9
            ---------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 112677 bases at least Q40
                Consensus quality: 121194 bases at least Q30
                Consensus quality: 125720 bases at least Q20
                Insert size: 129024; sum-of-contigs
                Quality coverage: 4.32x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            25 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     15833 contig of  15833 bp in length
              16334     34080 contig of  17747 bp in length
              34581     46531 contig of  11951 bp in length
              47032     56604 contig of   9573 bp in length
              57105     65754 contig of   8650 bp in length
              66255     71355 contig of   5101 bp in length
              71856     77672 contig of   5817 bp in length
              78173     81813 contig of   3641 bp in length
              82314     87385 contig of   5072 bp in length
              87886     92261 contig of   4376 bp in length
              92762     97060 contig of   4299 bp in length
              97561    102247 contig of   4687 bp in length
             102748    106863 contig of   4116 bp in length
             107364    109944 contig of   2581 bp in length
             110445    113790 contig of   3346 bp in length
             114291    117299 contig of   3009 bp in length
             117800    121336 contig of   3537 bp in length
             121837    125424 contig of   3588 bp in length
             125925    128032 contig of   2108 bp in length
```

```
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
------------- Project Information
    Center project name: L5770
    Center clone name: 131_K_9
--------------.
* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      918: contig of 918 bp in length
*                   gap of unknown length
*      919     1824: contig of 906 bp in length
*                   gap of unknown length
*     1825     2752: contig of 928 bp in length
*                   gap of unknown length
*     2753     3669: contig of 917 bp in length
*                   gap of unknown length
*     3670     4548: contig of 879 bp in length
*                   gap of unknown length
*     4549     5462: contig of 914 bp in length
*                   gap of unknown length
*     5463     6358: contig of 896 bp in length
*                   gap of unknown length
*     6359     7273: contig of 915 bp in length
*                   gap of unknown length
*     7274     8172: contig of 899 bp in length
*                   gap of unknown length
*     8173     9067: contig of 895 bp in length
*                   gap of unknown length
*     9068    10000: contig of 933 bp in length
*                   gap of unknown length
*    10001    10900: contig of 900 bp in length
*                   gap of unknown length
*    10903    11815: contig of 915 bp in length
*                   gap of unknown length
*    11816    12726: contig of 911 bp in length
*                   gap of unknown length
*    12727    13646: contig of 920 bp in length
*                   gap of unknown length
*    13647    14558: contig of 912 bp in length
*                   gap of unknown length
*    14559    15451: contig of 893 bp in length
*                   gap of unknown length
*    15452    16393: contig of 942 bp in length
*                   gap of unknown length
*    16394    17255: contig of 862 bp in length
*                   gap of unknown length
*    17256    18153: contig of 898 bp in length
*                   gap of unknown length
*    18154    19033: contig of 880 bp in length
*                   gap of unknown length
*    19034    19966: contig of 933 bp in length
*                   gap of unknown length
*    19967    20895: contig of 929 bp in length
*                   gap of unknown length
*    20896    21813: contig of 918 bp in length
*                   gap of unknown length
*    21814    22713: contig of 900 bp in length
*                   gap of unknown length
*    22714    23608: contig of 895 bp in length
*                   gap of unknown length
*    23609    24519: contig of 911 bp in length
*                   gap of unknown length
*    24520    25434: contig of 915 bp in length
*                   gap of unknown length
*    25435    26309: contig of 875 bp in length
*                   gap of unknown length
*    26310    27226: contig of 917 bp in length
*                   gap of unknown length
*    27227    28127: contig of 901 bp in length
*                   gap of unknown length
*    28128    29027: contig of 900 bp in length
*                   gap of unknown length
*    29028    29943: contig of 916 bp in length
*                   gap of unknown length
*    29944    30848: contig of 905 bp in length
*                   gap of unknown length
*    30849    31764: contig of 916 bp in length
*                   gap of unknown length
*    31765    32673: contig of 909 bp in length
*                   gap of unknown length
*    32674    33584: contig of 911 bp in length
*                   gap of unknown length
*    33585    34503: contig of 919 bp in length
*                   gap of unknown length
*    34504    35390: contig of 887 bp in length
*                   gap of unknown length
*    35391    36287: contig of 897 bp in length
*                   gap of unknown length
*    36288    37171: contig of 884 bp in length
*                   gap of unknown length
*    37172    38102: contig of 931 bp in length
*                   gap of unknown length
*    38103    38995: contig of 893 bp in length
*                   gap of unknown length
*    38996    39916: contig of 921 bp in length
*                   gap of unknown length
*    39917    40854: contig of 938 bp in length
*                   gap of unknown length
*    40855    41771: contig of 917 bp in length
*                   gap of unknown length
*    41772    42674: contig of 903 bp in length
*                   gap of unknown length
*    42675    43522: contig of 848 bp in length
*                   gap of unknown length
*    43523    44459: contig of 937 bp in length
*                   gap of unknown length
*    44460    45411: contig of 952 bp in length
*                   gap of unknown length
*    45412    46349: contig of 938 bp in length
*                   gap of unknown length
*    46350    47276: contig of 927 bp in length
*                   gap of unknown length
*    47277    48173: contig of 897 bp in length
*                   gap of unknown length
*    48174    49064: contig of 891 bp in length
*                   gap of unknown length
*    49065    49947: contig of 883 bp in length
*                   gap of unknown length
*    49948    50850: contig of 903 bp in length
*                   gap of unknown length
*    50851    51768: contig of 918 bp in length
*                   gap of unknown length
*    51769    52695: contig of 927 bp in length
*                   gap of unknown length
*    52696    53602: contig of 907 bp in length
*                   gap of unknown length
*    53603    54527: contig of 925 bp in length
*                   gap of unknown length
*    54528    55444: contig of 917 bp in length
*                   gap of unknown length
*    55445    56353: contig of 909 bp in length
*                   gap of unknown length
```

```
*    56354    57264: contig of 911 bp in length
*                 gap of unknown length
*    57265    58154: contig of 890 bp in length
*                 gap of unknown length
*    58155    59076: contig of 922 bp in length
*                 gap of unknown length
*    59077    60005: contig of 929 bp in length
*                 gap of unknown length
*    60006    60897: contig of 892 bp in length
*                 gap of unknown length
*    60898    61798: contig of 901 bp in length
*                 gap of unknown length
*    61799    62706: contig of 908 bp in length
*                 gap of unknown length
*    62707    63615: contig of 909 bp in length
*                 gap of unknown length
*    63616    64547: contig of 932 bp in length
*                 gap of unknown length

  Query Match             8.4%;  Score 18;  DB 51;  Length 83536;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    173 TTCTCTCTAATAAGAAAA 190
          ||||||||||||||||||
Db  60164 TTCTCTCTAATAAGAAAA 60147


RESULT  11
AL136089
LOCUS       AL136089   102995 bp    DNA              HTG       08-APR-2000
DEFINITION  Homo sapiens chromosome 6 clone RP1-278E11, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL136089
VERSION     AL136089.9  GI:7530184
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102995)
  AUTHORS   Smalley,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2000 this sequence version replaced gi:7320935.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00643  Length: 7736bp
            Contig_ID: 00773  Length: 17572bp
            Contig_ID: 00923  Length: 76087bp.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     7736: contig of 7736 bp in length
            *    7737 8536: gap of      800 bp
            *    8537    26108: contig of 17572 bp in length
            *   26109 26908: gap of      800 bp
            *   26909   102995: contig of 76087 bp in length.
FEATURES             Location/Qualifiers
     source          1. .102995
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP1-278E11"
                     /clone_lib="RPCI-1"
BASE COUNT    28381 a  22682 c  22914 g  27418 t   1600 others
ORIGIN


  Query Match             8.4%;  Score 18;  DB 40;  Length 102995;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 CTTTGAAACATCTACTGG 214
          ||||||||||||||||||
Db  77128 CTTTGAAACATCTACTGG 77145


RESULT  12
HS86F14/c
LOCUS       HS86F14    106571 bp    DNA              PRI       23-NOV-1999
DEFINITION  Human DNA sequence from PAC 86F14 on chromosome 1q23-1q24. Contains
            coagulation factor V, ESTs and STS.
ACCESSION   Z99572
VERSION     Z99572.1  GI:2769646
KEYWORDS    1q23-1q24; blood coagulation factor; factor V.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106571)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome 1 Project Group
            (http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 13, 1998 this sequence version replaced gi:2578147.
            IMPORTANT: This sequence is not the entire insert of clone 86F14.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre chromosome 1
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true right end of clone 206D15 is at 104.
            The true right end of clone 86F14 is at 106571.
            86F14 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .106571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q23-1q24"
                     /clone="RP1-86F14"
                     /clone_lib="RPCI-1"
     repeat_region   810. .1090
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   1270. .1360
                     /note="MLT1E repeat: matches 1. .84 of consensus"
```